GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: February 5, 2004, 13:18:44 ; Search time 46 Seconds
(without alignments)
750.534 Million cell updates/sec

Title: US-09-972-758A-2
Perfect score: 1910
Sequence: 1 MAEPFLSEYQHQPQTSNCTG.....LTENELHRQQERAPLSKFGD 359

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	148.5	7.8	992	2	T46337	hypothetical prote
2	139	7.3	755	2	S32103	filensin - bovine
3	134	7.0	568	1	I61106	involucrin - rat
4	133	7.0	462	1	A60746	chromogranin A pre
5	133	7.0	538	1	FOMV1M	gag polyprotein -
6	133	7.0	1493	2	A44224	DNA repair helicas
7	132.5	6.9	530	2	JC7168	lens epithelium-de
8	131	6.9	681	2	E59436	KIAA1314 protein [
9	131	6.9	1095	2	T00329	hypothetical prote
10	131	6.9	1898	1	A45973	trichohyalin - hum
11	131	6.9	3488	2	T34418	hypothetical prote
12	129	6.8	1054	2	T30177	cytoskeleton assem
13	128	6.7	1733	1	B45344	probable nuclear a
14	127.5	6.7	852	2	T06310	hypothetical prote
15	127.5	6.7	918	2	A88188	protein C18H9.3 [i
16	127.5	6.7	4687	1	A39638	plectin - rat
17	126.5	6.6	777	2	F54024	protein kinase (EC
18	126.5	6.6	1407	1	S28589	trichohyalin - rab
19	125.5	6.6	779	2	E54024	protein kinase (EC
20	125.5	6.6	1549	1	A40691	trichohyalin - she
21	125.5	6.6	1909	2	A45592	liver stage antige
22	125	6.5	383	2	C96581	hypothetical prote
23	125	6.5	913	2	T52485	neurofilament prot
24	124.5	6.5	474	2	T16441	hypothetical prote
25	123.5	6.5	677	1	S09078	chromogranin B pre
26	123.5	6.5	693	2	JC7925	nucleolin - common
27	123.5	6.5	1870	2	S37671	MHC class III hist
28	123.5	6.5	1872	2	S36152	MHC class III hist
29	123	6.4	723	2	A48217	single-strand DNA/
30	123	6.4	1829	2	T41751	l-afadin - rat
31	122	6.4	705	2	S32644	nucleolin - Africa
32	122	6.4	723	2	S33688	hypothetical prote
33	121.5	6.4	1974	2	T30010	hypothetical prote
34	121	6.3	538	1	FOMVM	gag polyprotein -
35	120.5	6.3	849	1	S64732	scaffold attachmen
36	120	6.3	536	1	FOMVME	gag polyprotein -
37	120	6.3	699	2	I38073	nucleolar phosphop
38	120	6.3	1325	2	S16129	dynein-associated
39	120	6.3	1663	2	T42092	s-afadin - rat
40	119.5	6.3	451	2	G70241	hypothetical prote
41	119.5	6.3	651	2	F86563	hypothetical prote
42	119.5	6.3	651	2	A72060	hypothetical prote
43	119.5	6.3	1233	2	T14157	serine/threonine p
44	119.5	6.3	1560	2	T30282	calcium-binding pr
45	119	6.2	449	1	A41520	chromogranin A pre

ALIGNMENTS

RESULT 1
T46337
hypothetical protein DKFZp434O2413.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T46337
R;Koehrer, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z23037
A;Accession: T46337
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-992 <AAA>
A;Cross-references: EMBL:AL137265
A;Experimental source: adult testis; clone DKFZp434O2413
C;Genetics:
A;Note: DKFZp434O2413.1

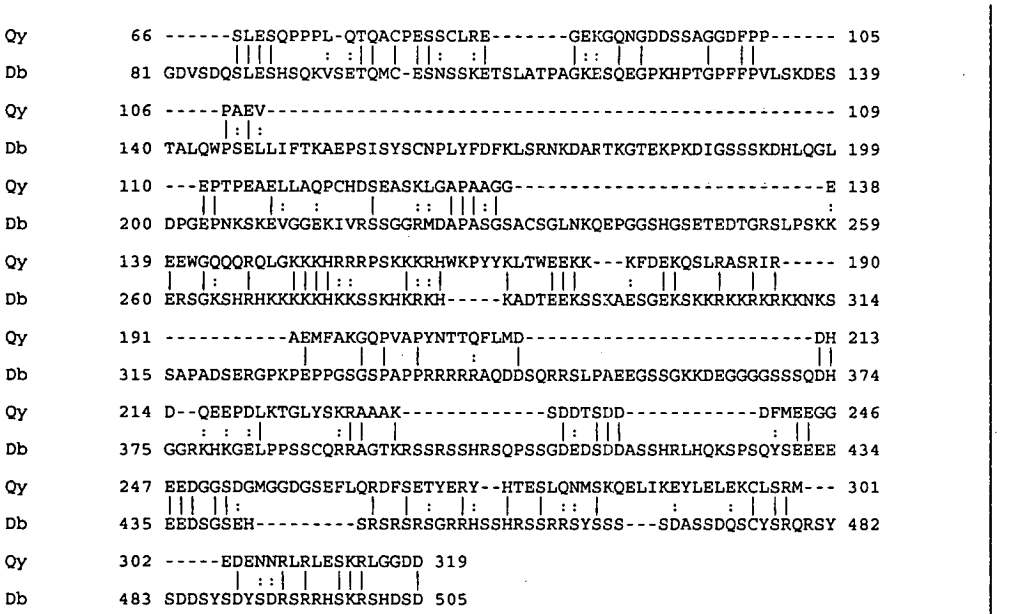
Query Match 7.8%; Score 148.5; DB 2; Length 992;
Best Local Similarity 21.1%; Pred. No. 0.19;
Matches 94; Conservative 66; Mismatches 149; Indels 137; Gaps 17;

```
Qy         25 QEELNPERPPGAEERVP-----EEDSRWQSRAF------PQLGGRPGPEGE--------- 64
              |:||  ::  | |||:     '|| ::   |:        ||     ||||:
Db        230 QDELQSKQSKGLEERLSPPLPHEERAQSPPRSLATEEEPPQ-----GPEGQPEWKEAEEL 284

Qy         65 -----GSLESQPPPLQTQA-CPESSCLREGEKGQNGDDSSAGGDFPPPAE----VEPTP- 113
                  ||  |    : ||  | ::| :  |:    ::   | :     |    | |||
Db        285 GEDSAASLSLQLSLQREQAPSPPAACEKGKEQHSQAEELGPGQEEAEDPEEKVAVSPTPP 344

Qy        114 ------EAELLAQPCHDSEASKLGAPAAGGEEEWGQQQRQLGKKK--------------- 152
                   | :| |   |||: | |       :   | || | : |
Db        345 VSPEVRSTEPVAPPEQLSEAA-LKAMEEAVAQVLEQDQRHLLESKQEKMQQLREKLCQEE 403

Qy        153 -------HRRRPSKKKRHWKPYYKLTWEEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNT 205
                    |:::       :   |   ||: :  |::| | | :||:: :  |
Db        404 EEEILRLHQQKEQSLSSLRERLQKAIEEEEARMREEESQRLSWLRAQVQSSTQA------ 457

Qy        206 TQFLMDDHDQEEPDLKTGLYSKRAAAKSD--------DTSDDDFMEEGGEEDGGS----- 252
                | ||   : :  |   |  :|         :  :  :|:  ||   |
Db        458 ------DEDQIRAEQEASLQKLREELESQQKAERASLEQKNRQMLEQLKEEIEASEKSEQ 511

Qy        253 -----------------DGMGGDGSEFLQRDFSETYERY-------HTE---SLQNMSK 284
                             :|   :    |::: |   ||         | |   |||   :
Db        512 AALNAAKEKALQQLREQLEGERKEAVATLEKEHSAELERLCSSLEAKHREVVSSLQKKIQ 571

Qy        285 QELIKEYLELEKCLSRMEDENNRLRLESKRLGGDDARVREL------------ELELDRL 332
              :  ||  :|:||| ::|   :|:  :|  : | :  :  |             |  ||::
Db        572 EAQQKEEAQLQKCLGQVE---HRVHQKSYHVAGYEHELSSLLREKRQEVEGEHERRLDKM 628

Qy        333 RAENLQLLT----ENELHRQQERAPL 354
              : |: |::     : |   :::|| |
```

```
Db        629 KEEHQQVMAKAREQYEAEERKQRAEL 654


RESULT 2
S32103
filensin - bovine
N;Alternate names: intermediate filament protein
C;Species: Bos primigenius taurus (cattle)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 10-Sep-1997
C;Accession: S32103; A40690
R;Gounari, F.; Merdes, A.; Quinlan, R.; Hess, J.; FitzGerald, P.G.; Ouzounis, C.; Georga
submitted to the EMBL Data Library, March 1993
A;Description: Bovine filensin possesses primary and secondary structure similarity to i
A;Reference number: S32103
A;Accession: S32103
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-755 <GOU>
A;Cross-references: EMBL:X72388; NID:g287751; PID:g287752
R;Gounari, F.; Merdes, A.; Quinlan, R.; Hess, J.; FitzGerald, P.G.; Ouzounis, C.A.; Geor
J. Cell Biol. 121, 847-853, 1993
A;Title: Bovine filensin possesses primary and secondary structure similarity to interme
A;Reference number: A40690; MUID:93260017; PMID:8491777
A;Accession: A40690
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-622,'RP',625-755 <GO2>
A;Experimental source: lens
A;Note: sequence extracted from NCBI backbone (NCBIN:132495, NCBIP:132499)
A;Note: part of this sequence was confirmed by protein sequencing
C;Keywords: membrane-associated protein

  Query Match             7.3%;  Score 139;  DB 2;  Length 755;
  Best Local Similarity   22.6%;  Pred. No. 0.51;
  Matches   88;  Conservative   32;  Mismatches  117;  Indels  152;  Gaps   16;

Qy         26 EELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPGPEGEGSLESQPP-PLQTQA------ 78
              || :|    || | ||  |    |:|| :||    : :|  | :||  | |:
Db        414 EEASPPTQEGAPEDVP--DGGKISKAFEKLGKMIKEKVKGPKEPEPPADLYTKGRYVMVS 471

Qy         79 ----------CPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTP--------EAELLAQ 120
                       |  |   : |     ||||      || : |||:|        | :   |
Db        472 GDGSFVDPGFCVFSVPAKGGVVVSKGDDS-----VPPDSGVEPSPQQPEPPLEEGQGPPQ 526

Qy        121 PCHDSEASKLGAPAAGGE---------------------------------------EEWG 142
                 |    : | |  ||                                         || |
Db        527 EKEDGLKEEGGPPEGKGEPPEGKGDSVKEEGGPPEGKGDGVKEEGGPPEGKGDGVKEEGG 586

Qy        143 QQQ-RQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDEKQSLRASRIRAEMFAKGQPVA 201
                : :  | ||    |  |    |       ||:    ||:               ||
Db        587 PPEGKGDGVKKEGEPPEGKGEGLK-------EEEGPLQEKED-------------GQSPT 626

Qy        202 PYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDDFMEEG----------GEED-- 249
                |:       |  |::      ||  |:          ||  |||          | |
Db        627 PHPA-----DKGDEKNAKELKGLQGKQ----------DDQKEEGARGPCPMVAPGPEGPS 671

Qy        250 -----------GGSDGMGGDGSEFLQR-----------DFSETYERYHTESLQNMSKQEL 287
                        |||:| |      | |           :  |: |:: |||:|   :  :
Db        672 TPRSQGPQVILGGSEGHGARSGSRLARSPPRKLAYEKVEVMESIEKFSTESIQTYEETAV 731

Qy        288 IKEYLELEKCLSRMEDENNRLRLESKRLG 316
                | | : :||           :   |:||
Db        732 IVETM-IEK----------TKANKKKLG 748


RESULT 3
I61106
involucrin - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 02-Aug-1996 #sequence_revision 23-Aug-1996 #text_change 22-Jun-1999
C;Accession: I61106
R;Djian, P.; Phillips, M.; Easley, K.; Huang, E.; Simon, M.; Rice, R.H.; Green, H.
Mol. Biol. Evol. 10, 1136-1149, 1993
A;Title: The involucrin genes of the mouse and the rat: study of their shared repeats.
A;Reference number: A49377; MUID:94104476; PMID:8277848
A;Accession: I61106
A;Status: translation not shown; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-568 <RES>
A;Cross-references: GB:L28818; NID:g454427; PIDN:AAA41445.1; PID:g454428
C;Comment: During the terminal differentiation of keratinocytes, this protein from the cy
linked envelope under the plasma membrane.
C;Superfamily: involucrin
C;Keywords: cornified cell envelope; duplication; epidermis; tandem repeat

  Query Match             7.0%;  Score 134;  DB 1;  Length 568;
  Best Local Similarity   20.9%;  Pred. No. 0.76;
  Matches   93;  Conservative   66;  Mismatches  169;  Indels  118;  Gaps   18;

Qy          3 EPFLSEYQHQPQTSNCTGAAAVQEE-----------------LNPERPPGAEERVPEED 44
              | | ::|   | |   |   |::                   || | | ::: |||
Db        127 EQHLRQHQQPQQESQGQGLCLGQQQDVLAPQELHMGQHQKEKLQEPELPLGQQQKTPEEQ 186

Qy         45 -----SRWQSRAFPQLGGRPGPE----GEGSLESQPPPLQTQAC---------PESSCLR 86
                   : |     :   | :    |:  : ||  : |           || |||
Db        187 ELILGEKQQKLHLVERHQEPQEQELHHGQKQKQQQPQEQELQLVQHQKQKQHEPE-LCLR 245

Qy         87 EGEKGQNGDDSSAGGDFPPPAEVEP------------TPEAELLAQPCHDSEASKLGAP 133
              : :: :: :    |      ||              ||  |  |   :|   :|
Db        246 KQQQQESHERELHLGKQQQQESHEPELHLGKQQHQESHEPELHLGKQQHQESCEPEL--- 302

Qy        134 AAGGEEEWGQQQRQ--------LGKKKHRR------RPSKKKRHWKPYYKLTWEEKKKF- 178
                | ::   ||| |        |||::|:       :  |:::  :|   |  :||:|
Db        303 HLGEQQHQEQQQHQESCEPELHLGKQQHQETQESELQLGKQQKPHEPDMVLDPKEKQKLH 362

Qy        179 --------DEKQSLRASRIRAEM------------FAKGQPVAPYNTTQFLMDDHDQEEP 218
                     : |  : | :: |                | | :   :  |:   :     ||
Db        363 DPELHLGKQQHQESQESELQVEKKQHEKSPEPELHLGKQQELHEPDMTEDQKEKQSLHEP 422

Qy        219 DLKTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTES 278
              :|  |      |  :: : |  |:  |:     :      | :  |:   | |:|  ::|
Db        423 ELHLG--------KQQESHEPDMTEDQKEKQSLYEPELHLGKQQEQQIEYEGYQR--SKS 472

Qy        279 LQNMSKQELIKEYLELEKCLSRMEDENNRL--RLESKRLGGDDARVRELELELDRLRAEN 336
              |  : |||       ||:   | :| |   |  ||: : :  |:       :||  : : ||
Db        473 LNQLLKQEKASRGQELDD--SHLEQEKELLDQRLDQELVNKDE------QLERKKHKLEN 524

Qy        337 L--------QLLTENELHRQQERAPL 354
              |        ||:   :  | ||  |:
Db        525 LTQKEKQIKQLVPSTD--RVQETQPI 548


RESULT 4
A60746
chromogranin A precursor - rat
N;Contains: beta-granin; pancreastatin
C;Species: Rattus norvegicus (Norway rat)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C;Accession: A60746; S00291; S02543
R;Parmer, R.J.; Koop, A.H.; Handa, M.T.; O'Connor, D.T.
Hypertension 14, 435-444, 1989
A;Title: Molecular cloning of chromogranin A from rat pheochromocytoma cells.
A;Reference number: A60746; MUID:90007662; PMID:2793216
A;Accession: A60746
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-462 <PAR>
A;Cross-references: GB:AF145445; NID:g5163367; PIDN:AAD40652.1; PID:g5163368
R;Iacangelo, A.; Okayama, H.; Eiden, L.E.
FEBS Lett. 227, 115-121, 1988
A;Title: Primary structure of rat chromogranin A and distribution of its mRNA.
A;Reference number: S00291; MUID:88112232; PMID:2828116
A;Accession: S00291
A;Molecule type: mRNA
A;Residues: 1-107,'QQQQ',108-462 <IAC>
A;Cross-references: EMBL:X06832; NID:g55950; PIDN:CAA29988.1; PID:g55951
```

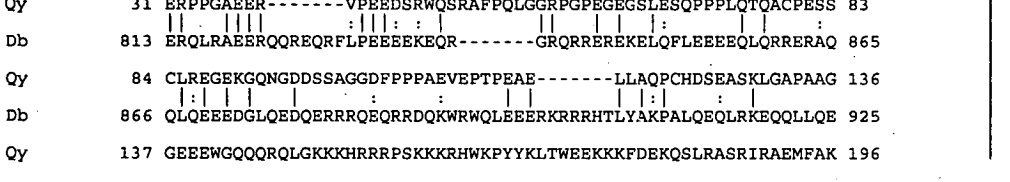
```
R;Hutton, J.C.; Nielsen, E.; Kastern, W.
FEBS Lett. 236, 269-274, 1988
A;Title: The molecular cloning of the chromogranin A-like precursor of beta-granin and p
A;Reference number: S02543; MUID:88312980; PMID:3044825
A;Accession: S02543
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 19-107,'QQQQ',108-462 <HUT>
A;Note: part of this sequence, including the amino end of beta-granin, was determined by
C;Superfamily: chromogranin A
C;Keywords: glycoprotein; phosphoprotein
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-462/Product: chromogranin A #status predicted <MAT>
F;92-107/Region: glutamine-rich
F;417-419/Region: cell attachment (R-G-D) motif
F;35-56/Disulfide bonds: #status predicted
F;68,147,216,349,358,386,408,409,413/Binding site: phosphate (Ser) (covalent) #status pr
F;189/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             7.0%;  Score 133;  DB 1;  Length 462;
  Best Local Similarity   24.2%;  Pred. No. 0.7;
  Matches   87;  Conservative   54;  Mismatches  159;  Indels   60;  Gaps   17;

Qy          3 EPFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPGPE 62
              |  ||  :||          : :||      ||:||  ::  : | :   |       |
Db         64 ERVLSILRHQ----------NLLKELQDLALQGAKERAQQQQQQQQQQQ-QQQQQHSSFE 112

Qy         63 GEGS--LESQPPPLQ----TQACPESSCLREGEKGQNGDDSSAGG--DFP-PPAEVEPTP 113
              | |   |:| |  :       |    : : |    |   :  |  : | | |  ||
Db        113 DELSEVFENQSPAAKHGDAASEAPSKDTVEKREDSDKGQQDAFEGTTEGPRPQAFPEPKQ 172

Qy        114 EAELL--AQPCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPS--KKKRHWKPYYK 169
              |: ::  :|   :  |:  :| :  :| | |  | :  | ||  :: | |  |
Db        173 ESSMMGNSQSPGEDTANNTQSPTSLPSQEHGIPQTTEGSE---RGPSAQQQARKAKQEEK 229

Qy        170 LTWEEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNT------TQFLMDDHDQEEPDLKTG 223
              ||:|: :|::   :  | :  |  | |  ::      :   ||  | |       |
Db        230 EEEEEEKEEEEEEKEEKAIAREKAGPKEVPTAASSSHFYSGYKKIQKDDDGQSESQAVNG 289

Qy        224 LYSKRAAAKSDDTSDDDFMEEGGEEDGGSDGM---------GGDGSEFLQRDFSETYERY 274
              |  |::: :    :|  :|: | : |          || | | |:|   |  |
Db        290 ---KTGASEAVPSEGKGELEHSQQEEDGEEAMAGPPQGLFPGGKGQE-LERKQQEEEE-- 343

Qy        275 HTESLQNMSKQELIKEYLELEKCLSRMEDENNRLRLESKRLGGDDARVRELELELDRLRA 334
                  :: | :|:   :|  |||:     |  | ||| |:|    | ::|   | ||
Db        344 --------EEERLSREW--EDKRWSRMDQLAKELTAE-KRLEGEDDPDRSMKLSF-RARA 391


RESULT 5
FOMV1M
gag polyprotein - Moloney murine leukemia virus
N;Contains: core protein p15; core shell protein p30; inner coat protein p12; nucleoprot
C;Species: Moloney murine leukemia virus
A;Note: host Mus sp. (mouse)
C;Date: 01-Sep-1981 #sequence_revision 27-Nov-1985 #text_change 24-Jul-1997
C;Accession: A03930
R;Shinnick, T.M.; Lerner, R.A.; Sutcliffe, J.G.
Nature 293, 543-548, 1981
A;Title: Nucleotide sequence of Moloney murine leukaemia virus.
A;Reference number: A93265; MUID:82035843; PMID:6169994
A;Accession: A03930
A;Molecule type: genomic RNA
A;Residues: 1-538 <SHI>
A;Experimental source: clone pMLV-1
C;Comment: This protein is synthesized as a gag-pol polyprotein.
C;Genetics:
A;Gene: gag
C;Superfamily: mammalian retrovirus gag polyprotein I
C;Keywords: core protein; inner coat protein; nucleoprotein; polyprotein
F;2-131/Product: core protein p15 #status predicted <C15>
F;132-215/Product: inner coat protein p12 #status predicted <C12>
F;216-478/Product: core shell protein p30 #status predicted <C30>
F;479-534/Product: nucleoprotein p10 #status predicted <C10>

  Query Match             7.0%;  Score 133;  DB 1;  Length 538;
  Best Local Similarity   20.9%;  Pred. No. 0.82;
  Matches   89;  Conservative   40;  Mismatches  142;  Indels  154;  Gaps   19;

Qy         30 PERPPGA-------EERVPEEDSRWQSRAFPQLGGRPGPE---GEGS-----LESQPPPL 74
              |  || |           |   | : :   | || :| |:      |       |  |||
Db        107 PPLPPSAPSLPLEPPRSTPPRSSLYPALT-PSLGAKPKPQVLSDSGGPLIDLLTEDPPPY 165

Qy         75 QTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELL----AQPCHDSEASKL 130
              :   | |   |:|    || :::  |: |   :|:| | |      |  ||  |:
Db        166 RDPRPPPSD--RDG----NGGEATPAGEAP-----DPSPMASRLRGRREPPVADSTTSQA 214

Qy        131 GAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDEKQSLRASRIR 190
                 |||           |: ::    |    :||          | |       : |
Db        215 FPLRAGGN----------GQLQYWPFSSSDLYNWK-------NNNPSFSEDPGKLTALIE 257

Qy        191 AEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTG------LYSKRAAAKSDD--------- 235
              : :           | |   ||  |    | ||       |   | | : ||
Db        258 SVLI----------THQPTWDDCQQLLGTLLTGEEKQRVLLEARKAVRGDDGRPTQLPNE 307

Qy        236 --------TSDDDFMEEGGEE-------------------------DGGSDGMGGDGSE 261
                     | |:  : |                           | : |    |
Db        308 VDAAFPLERPDWDYTTQAGRNHLVHYRQLLLAGLQNAGRSPTNLAKVKGITQGPNESPSA 367

Qy        262 FLQRDFSETYERY-----------------------------HTESLQNMSKQELIKE 290
              ||:|   | | ||                           | |:| :  :|::
Db        368 FLER-LKEAYRRYTPYDPEDPGQETNVSMSFIWQSAPDIGRKLGRLEDLKNKTLGDLVR- 425

Qy        291 YLELEKCLSRMEDENNRLRLESKRLGGDDARV-RELELELDRLRAENLQLLTENELHRQQ 349
              | ||  :: |   |          : |: || | : :| | | |: |   | :  | :
Db        426 --EAEKIFNKRETPEER----------EERIRRETEEKEERRRTEDEQKEKERDRRRHR 472

Qy        350 ERAPL 354
              | : |
Db        473 EMSKL 477


RESULT 6
A44224
DNA repair helicase ERCC6 - human
C;Species: Homo sapiens (man)
C;Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C;Accession: A44224
R;Troelstra, C.; van Gool, A.; de Wit, J.; Vermeulen, W.; Bootsma, D.; Hoeijmakers, J.H.
Cell 71, 939-953, 1992
A;Title: ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's
A;Reference number: A44224; MUID:93092214; PMID:1339317
A;Accession: A44224
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-1493 <TRO>
A;Cross-references: GB:L04791; NID:g182180; PIDN:AAA52397.1; PID:g182181
A;Note: sequence extracted from NCBI backbone (NCBIP:120238)
C;Keywords: DNA binding

  Query Match             7.0%;  Score 133;  DB 2;  Length 1493;
  Best Local Similarity   23.7%;  Pred. No. 2.3;
  Matches   82;  Conservative   42;  Mismatches  118;  Indels  104;  Gaps   17;

Qy         35 GAEERVP------EEDSRWQSRAFPQLGGRPGPEGEGSLESQPPPLQTQACPESSCLREG 88
              ||| ::       |||:              |||    || |   |:|  |  |   :| |
Db        201 GAEVKIELDHASLEEDA-----------EPGP---SSLGSMLMPVQETAWEE--LIRTG 243

Qy         89 EKGQNGDDSSAGGDFPPPAEVEPTP----EA--------------ELLAQPCHDSEASK 129
              :       :  |   |   | :|      ||                |   | |:   | |
Db        244 QM------TPFGTQIPQKQEKKPRKIMLNEASGFEKYLADQAKLSFERKKQGCNKRAARK 297

Qy        130 LGAPA-----AGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDEKQSL 184
              ||          : :  :: | | ||: |      |:| |   |   : : |    ::
Db        298 APAPVTPPAPVQNKNKPNKKARVLSKKEER-----LKKHIKKLQKRALQFQGKVGLPKAR 352

Qy        185 RA----SRIRAEMFAKGQPVAPYNTTQFLMDDHDQE----EPDL-------------KTG 223
```

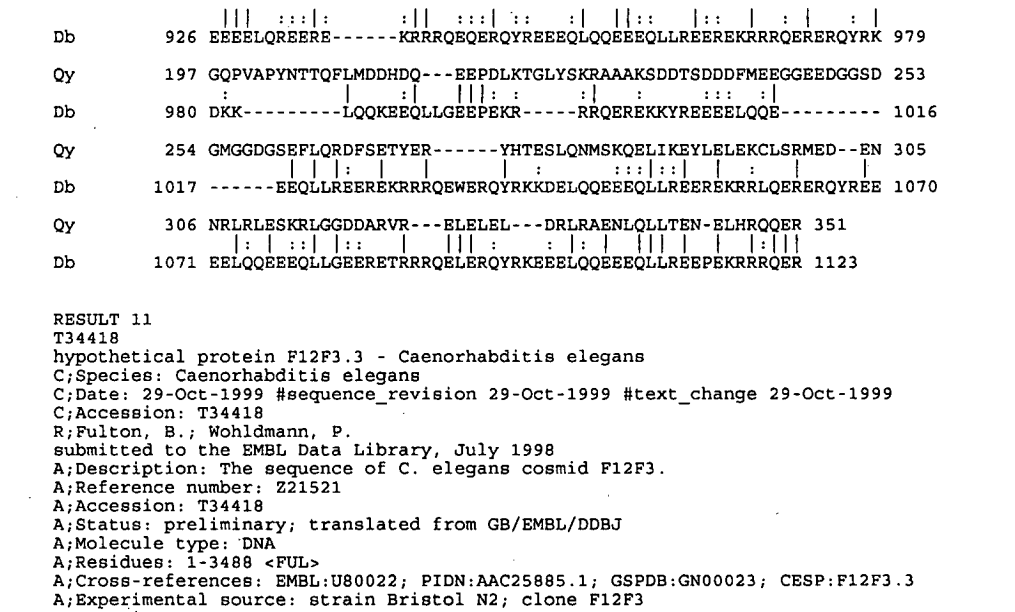
```
Db        353 RPWESDMRPEAEGDSEGEE-SEYFPTEEEEEEEDDEVEGAEADLSGDGTDYELKPLPKGG 411

Qy        224 LYSKRAAAKSDDTSDDDFMEEGGEE-DGGSDGMGGDGSEFLQRDFSETYERYHTESLQNM 282

Db        412 KRQKKVPVQE---IDDDFFPSSGEEAEAASVGEGGGGGRKVGRYRDDGDEDYYKQRLRRW 468

Qy        283 SKQELIKEYLELEKCLSRMEDENNRLRLESKRLGGDDARVRELELE 328

Db        469 NK--------------LRLQDKEKRLKLE------DDSEESDAEFD 494


RESULT 7
JC7168
lens epithelium-derived growth factor - human
N;Alternate names: transcriptional co-activator p75 protein
C;Species: Homo sapiens (man)
C;Date: 04-Mar-2000 #sequence_revision 04-Mar-2000 #text_change 11-May-2000
C;Accession: JC7168
R;Singh, D.P.; Ohguro, N.; Kikuchi, T.; Sueno, T.; Reddy, V.N.; Yuge, K.; Chylack Jr., L
Biochem. Biophys. Res. Commun. 267, 373-381, 2000
A;Title: Lens epithelium-derived growth factor: Effects on growth and survival of lens e
A;Reference number: JC7168; MUID:20090645; PMID:10623627
A;Accession: JC7168
A;Molecule type: mRNA
A;Residues: 1-530 <SIN>
A;Cross-references: GB:AF063020
A;Experimental source: lens epithelial cells
C;Comment: This protein, a regulatory factor, belongs to a family of growth factor, whic
on of this factor stimulates cell attachment, shortenes the initial lag in growth, and e
C;Genetics:
A;Gene: ledgf
C;Keywords: eye lens; glycoprotein; growth factor; growth regulation

  Query Match             6.9%;  Score 132.5;  DB 2;  Length 530;
  Best Local Similarity   21.2%;  Pred. No. 0.87;
  Matches   77;  Conservative   56;  Mismatches  115;  Indels  115;  Gaps   17;

Qy          7 SEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQ----------L 55

Db         93 SSQQAATKQSNASSDVEVEEK---------ETSVSKEDTDHEEKASNEDVTKAVDITTPK 143

Qy         56 GGRPG--PEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEV---E 110

Db        144 AARRGRKRKAEKQVETEEAGVVTTATASVNLKVSPKRGR----------PAATEVKIPK 192

Qy        111 PTPEAELLAQPCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKL 170

Db        193 PRGRPKMVKQPC-PSESDII------------TEEDKSKKKGQEEKQPKK--------- 229

Qy        171 TWEEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTG---LYSK 227

Db        230 ---QPKKDEEGQK-------------------------EEDKPRKEPDKKEGKKEVESK 260

Qy        228 RA-AAKSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSETYER------YHTESLQ 280

Db        261 RKNLAKTGVTSTSDSEEEGDDQEGEKKRKGG-------RNFQTAHRRNRLKGQHEKEAAD 313

Qy        281 NMSKQELIKEYLELEKCLSRMEDENNR---LRLESKRLGGDDARVRELELEL-DRLRAEN 336

Db        314 RKRKQE---EQMETEQ---QNKDEGKKPEVKKVEKKRETSMDSRLQRIHAEIKNSLKIDN 367

Qy        337 LQL 339

Db        368 LDV 370


RESULT 8
E59436
KIAA1314 protein [imported] - human
C;Species: Homo sapiens (man)
C;Date: 03-Jun-2002 #sequence_revision 03-Jun-2002 #text_change 23-Sep-2002
C;Accession: E59436; F59436
R;Nagase, T.; Kikuno, R.; Ishikawa, K.I.; Hirosawa, M.; Ohara, O.
DNA Res. 7, 65-73, 2000
A;Title: Prediction of the coding sequences of unidentified human genes. XVI. The complet
A;Reference number: E59436
A;Accession: E59436
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-681 <NAG>
A;Cross-references: GB:BAA92552; PID:g7243009; PIDN:BAA92552.1
R;Ohara, O.; Nagase, T.; Kikuno, R.
submitted to GenBank, January 2000
A;Reference number: F59436
A;Accession: F59436
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-681 <NAG>
A;Cross-references: GB:BAA92552; PID:g7243009; PIDN:BAA92552.1

  Query Match             6.9%;  Score 131;  DB 2;  Length 681;
  Best Local Similarity   26.4%;  Pred. No. 1.4;
  Matches   84;  Conservative   44;  Mismatches  110;  Indels   80;  Gaps   18;

Qy         66 SLESQPPPLQTQACPESSC--------LREGEKGQNGDDSSAGG-DFPPPAEVEPTPEAE 116

Db         14 SNESLHPPAFSRSNSEASVDSASMEDFWREIESIK---DSSMGGQEEPPPAEVTPVDEGE 70

Qy        117 LLAQPCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKK 176

Db         71 LEAEWLQDVGLSTL---ISGDEEEDG--KALLSTLTRTQAAAVQKR----YHTYTQTMRK 121

Qy        177 KFDEKQSLRASRIRAEMFAKGQ----PVAPYNTTQFLMDDHDQEEPDL-----KTGLYSK 227

Db        122 K--DKQSIRDVR---DIFGVSESPPRDTCGNHTNQL---DGTKEERELPRVIKTSGSMPD 173

Qy        228 RAAAKSDDTSDDDFMEEGGEEDGGSDGMG----------GDGS--------------EF 262

Db        174 DASLNSTTLSDASQDKEGSFAVPRSDSVAILETIPVLPVHSNGSPEPGQPVQNAISDDDF 233

Qy        263 LQRD---------FSETYERYHTESLQ--NMSKQELIKEYLELEKCLSRMEDENNRLRL- 310

Db        234 LEKNIPPEAEELSFEVSYSEMVTEALKRNKLKKSEIKKE----DYVLTKFNVQKTRFGLT 289

Qy        311 ESKRLGGDD-ARVRELEL 327

Db        290 EAGDLSAEDMKKIRHLSL 307


RESULT 9
T00329
hypothetical protein KIAA0553 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C;Accession: T00329
R;Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A;Title: Prediction of the coding sequences of unidentified human genes. IX. The complete
A;Reference number: Z14086; MUID:98290545; PMID:9628581
A;Accession: T00329
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1095 <NAG>
A;Cross-references: EMBL:AB011125; NID:g3043629; PIDN:BAA25479.1; PID:g3043630
A;Experimental source: brain
C;Genetics:
A;Note: KIAA0553

  Query Match             6.9%;  Score 131;  DB 2;  Length 1095;
  Best Local Similarity   19.7%;  Pred. No. 2.2;
  Matches   99;  Conservative   41;  Mismatches  137;  Indels  226;  Gaps   21;

Qy         13 PQTSNCTGAAAVQEELNPERPPGAEERVPE--EDSRWQSRAFPQLGGRPGPEGEG----- 65

Db         33 PKPKSCIKAAASQ---------GAEKTVSEVSEQPKETSMTEP---SEPGSKAEAKKALG 80
```



RESULT 10
A45973
trichohyalin - human
C;Species: Homo sapiens (man)
C;Date: 03-May-1994 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999
C;Accession: A45973
R;Lee, S.C.; Kim, I.G.; Marekov, L.N.; O'Keefe, E.J.; Parry, D.A.D.; Steinert, P.M.
J. Biol. Chem. 268, 12164-12176, 1993
A;Title: The structure of human trichohyalin. Potential multiple roles as a functional E
ed (cross-linking) protein.
A;Reference number: A45973; MUID:93280194; PMID:7685034
A;Accession: A45973
A;Molecule type: DNA
A;Residues: 1-1898 <LEE>
A;Cross-references: GB:L09190; NID:g292835; PIDN:AAA65582.1; PID:g292836
A;Note: authors translated the codon AGG for residue 1714 as Pro
C;Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sh
Covalent modifications to this protein include conversion of arginine to citrulline and
C;Genetics:
A;Gene: GDB:THH
A;Cross-references: GDB:136223; OMIM:190370
A;Map position: 1q21-1q21
C;Superfamily: trichohyalin; calmodulin repeat homology
C;Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
F;49-81/Domain: calmodulin repeat homology <EF2>

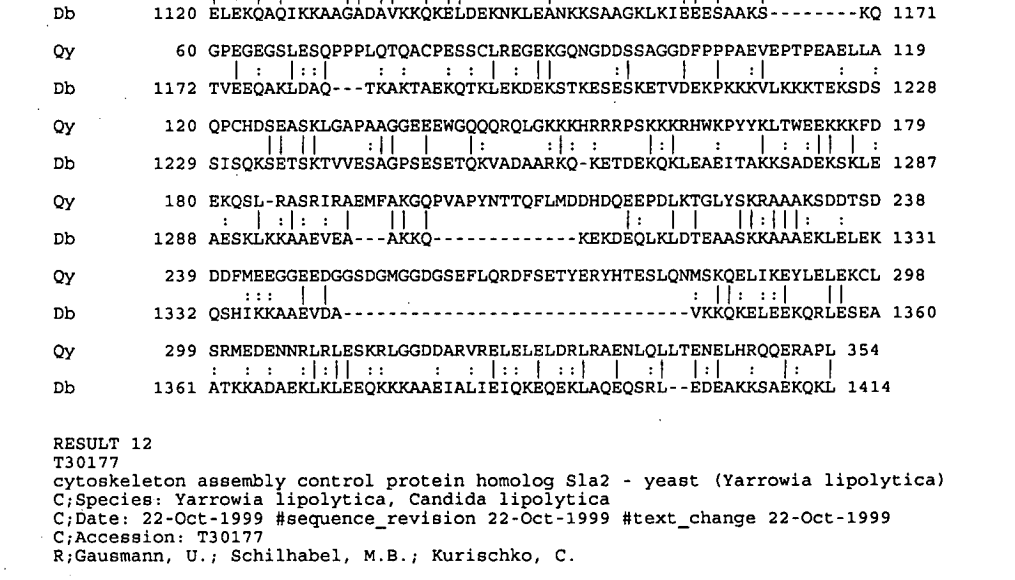
Query Match 6.9%; Score 131; DB 1; Length 1898;
Best Local Similarity 22.4%; Pred. No. 3.9;
Matches 79; Conservative 54; Mismatches 146; Indels 74; Gaps 13;





RESULT 11
T34418
hypothetical protein F12F3.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34418
R;Fulton, B.; Wohldmann, P.
submitted to the EMBL Data Library, July 1998
A;Description: The sequence of C. elegans cosmid F12F3.
A;Reference number: Z21521
A;Accession: T34418
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3488 <FUL>
A;Cross-references: EMBL:U80022; PIDN:AAC25885.1; GSPDB:GN00023; CESP:F12F3.3
A;Experimental source: strain Bristol N2; clone F12F3
C;Genetics:
A;Gene: CESP:F12F3.3
A;Map position: 5
A;Introns: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1

Query Match 6.9%; Score 131; DB 2; Length 3488;
Best Local Similarity 20.5%; Pred. No. 7.3;
Matches 73; Conservative 63; Mismatches 150; Indels 70; Gaps 10;



RESULT 12
T30177
cytoskeleton assembly control protein homolog Sla2 - yeast (Yarrowia lipolytica)
C;Species: Yarrowia lipolytica, Candida lipolytica
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C;Accession: T30177
R;Gausmann, U.; Schilhabel, M.B.; Kurischko, C.

```
submitted to the EMBL Data Library, July 1996
A;Description: Sla2 homologue of Yarrowia lipolytica.
A;Reference number: Z20763
A;Accession: T30177
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1054 <GAU>
A;Cross-references: EMBL:U65409; NID:g3978133; PID:g3978134; PIDN:AAC83184.1
A;Experimental source: strain W29
C;Genetics:
A;Gene: SLA2
A;Map position: III, adjacent to MATA

  Query Match             6.8%;  Score 129; DB 2;  Length 1054;
  Best Local Similarity   23.4%;  Pred. No. 2.8;
  Matches   92;  Conservative   45;  Mismatches  124;  Indels  132;  Gaps   19;

Qy         31 ERPPGAEERVPEEDSRWQSRAFPQLGGRPGPEGE-GSLESQPPPLQTQACPESSCLREGE 89
Db        262 KNPPNLLE---EEDG-------PGLPQRPKSVAETRSVSPEPAPLPV-ATPTPSI----- 305

Qy         90 KGQNGDDSSAGGDFPPPAEVEPT----PEAELLAQPCHDSEASKLGAPAAGGEEEWGQ-- 143
Db        306 ----------------PAEAQPIENFWSQDALLAQQQYDAEQERLRQQQAMEEERIRQMQ 349

Qy        144 -QQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDE----KQSLRASRIRAEMFAKGQ 198
Db        350 MQQQQQFEMQQRQQMEAQQRA---------QEQLMADQMARHAQGRMAELERDILALRGQ 400

Qy        199 PVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSDGMGGD 258
Db        401 ----YDQDQLMLEQYDGRVKALEAELNQLQQTAHQSAQAKDDLIES-------------- 442

Qy        259 GSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCL----------------SRM 301
Db        443 ----LQQQITMWRQKYETLAKRYSSMRE---EYLALLKKLKATQQKAASAKEAIEKAEKM 495

Qy        302 E--------------DENNRLRLESKRL-GGDDARVRELELEL----------DRLRAEN 336
Db        496 ERDMRHKNIELADLIKERDRARYDLDRAKGGNKEDVERLERELRMAQDKLADKDRSTGAD 555

Qy        337 LQLL----------TENELHRQQERAPLSKFGD 359
Db        556 LSLLLSKHNRELSELENAL-KMKQRA-LDERGD 586


RESULT 13
B45344
probable nuclear antigen - suid herpesvirus 1 (strain Kaplan)
C;Species: suid herpesvirus 1
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999
C;Accession: B45344
R;Vlcek, C.; Kozmik, Z.; Paces, V.; Schirm, S.; Schwyzer, M.
Virology 179, 365-377, 1990
A;Title: Pseudorabies virus immediate-early gene overlaps with an oppositely oriented op
A;Reference number: A45344; MUID:91021039; PMID:2171211
A;Accession: B45344
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-1733 <VLC>
A;Cross-references: GB:M34651; NID:g334070; PIDN:AAA47471.1; PID:g334072
C;Superfamily: pseudorabies virus 1 nuclear antigen

  Query Match             6.7%;  Score 128;  DB 1;  Length 1733;
  Best Local Similarity   22.4%;  Pred. No. 5.3;
  Matches   74;  Conservative   27;  Mismatches   94;  Indels  136;  Gaps   16;

Qy          8 EYQHQPQTSNCT--------------GAAAVQEE----LNPER-PPGAEER-------- 39
Db        104 DHQHRPPPTTTTTTIKDPQHPQDPLLLPTKTLQEEDPHLLRPTRDPPSAKTHHHQDPPGG 163

Qy         40 VPEEDSRWQSRAFPQLGGRPGP----------EGEGSLESQPPPLQ-------------- 75
Db        164 GPPSTSSHHHHQDPPGGGPPSPPPRPSTSSSSSHQGPPSTRPPPPQRPPPRWPPPSPQKI 223

Qy         76 --TQACPESSCLR------------EGEKGQNGDDSSAGG-----DFPPPAEVEPTPEA 115
Db        224 SETRAGSENTAQTLFSHSENKLFSHPMGEGGE-GDRGTAGGGEGDRDDPPPPSPPPRPPP 282

Qy        116 ELLAQPCHDSEASKLGAPAAGGEEEWGQQQRQLG---KKKHRRRPSKKKRHWKPYYKLTW 172
Db        283 PLPPPPPPPPPPQ---PPPAGGS---ARRRRGGGPPGRGGRRRGGKRRR---------- 326

Qy        173 EEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAK 232
Db        327 ----------------------AEGTEAA-------AADAEEEEDGD------------ 344

Qy        233 SDDTSDDDFMEEGGEEDGGSD--GMGGDGSE 261
Db        345 GDEDEDEDRAEGEGREDGGEGPRGAGGGAGE 375


RESULT 14
T06310
hypothetical protein F11C18.80 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999
C;Accession: T06310
R;Bevan, M.; Terryn, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.; De H
ewes, H.W.; Mayer, K.F.X.; Schueller, C.
submitted to the Protein Sequence Database, April 1999
A;Reference number: Z15589
A;Accession: T06310
A;Molecule type: DNA
A;Residues: 1-852 <BEV>
A;Cross-references: EMBL:AL049607; GSPDB:GN00062; ATSP:F11C18.80
A;Experimental source: cultivar Columbia; BAC clone F11C18
C;Genetics:
A;Gene: ATSP:F11C18.80
A;Map position: 4
A;Introns: 34/3; 102/3; 163/2; 205/3; 269/3; 282/3; 564/3; 602/3; 622/2; 639/3; 667/3; 68

  Query Match             6.7%;  Score 127.5;  DB 2;  Length 852;
  Best Local Similarity   23.2%;  Pred. No. 2.8;
  Matches   88;  Conservative   51;  Mismatches  145;  Indels   95;  Gaps   18;

Qy          1 MAEPPFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPG 60
Db        439 MAEE-LPEQSVPKKTANQKKKESSTEEVKPSASIATEE-VSEEPNTSE----PQVTKKSG 492

Qy         61 PEGEGSLESQPP-PLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEP----TPEA 115
Db        493 KKVASSSKTKPTVPPSKKSTSETKVAKQSEKKVVGSDNAQESTKPKEEKKKPGRGKAIDE 552

Qy        116 ELLAQPCHDSE--ASKLGAPAAGGEEEWGQ-----------QQRQLGKKKHRRRP---SK 159
Db        553 ESLHTSSGDNEKPAVSSGKLASKSKKEAKQTVEESPNSNTKRKRSLGQGKASGESLVGSR 612

Qy        160 KKRHW---KPYYKLTWEE----KKKF-----DEKQSLRASRIRAEMFAKGQPVAPYNTTQ 207
Db        613 IKVWWPMDQAYYKGVVESYDAAKKKHLVIYDDGDQEI--------LYLKNQKWSPLDESE 664

Qy        208 FLMD----DHDQEEPDLKT---GLYSKRA----AAKSDDTSDDD---------------- 240
Db        665 LSQDEEAADQTGQEEDASTVGSGAGSSKAKATPASKSSKTSQDDKTASKSKDSKEASREE 724

Qy        241 --FMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCL 298
Db        725 EASSEEESEEEEPPKTVGKSGSSRSKKDIS---------SVSKSGK----------SKAS 765

Qy        299 SRMEDENNRLRLESKRLGG 317
Db        766 SKKKEEPSKATTSSKSKSG 784


RESULT 15
```

```
A88188
protein C18H9.3 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C;Accession: A88188
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: A88188
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-918 <STO>
A;Cross-references: GB:chr_II; PID:g722383; GSPDB:GN00020; CESP:C18H9.3
A;Note: weakly similar to trichohyalin
C;Genetics:
A;Gene: C18H9.3
A;Map position: 2

  Query Match             6.7%;  Score 127.5;  DB 2;  Length 918;
  Best Local Similarity   21.0%;  Pred. No. 3;
  Matches   87;  Conservative   62;  Mismatches  150;  Indels  115;  Gaps   18;

Qy          4 PFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGG----RP 59
              |   : | | ::|     | : ||  : : |        | : | |: : : :||    :|
Db        307 PPQQQQQQQQRSSAPVSAPSRQESESTDVPNLPIPTYPSDPSAWSNNS---MGGGIFYQP 363

Qy         60 GPEGEGSLESQPP-------PLQTQACP-----------------ESSCLREGEKGQN-- 93
               |:     :  : ]       | :|:  |                 ||  :: || |:
Db        364 TPQPPAPIVKEEPVQFYYMDPTETRRGPFPKDQMNVWFKAGYFTDESLRVQRGENGEYKT 423

Qy         94 -GDDSSAGGDFPPPAEVEP-TPEAELLAQPCHDSEASKLGAPAAGGEEEW---GQ----- 143
               ||      |   |   :|   |   :|    : |  : |   | ||   |   ||
Db        424 IGDLKKLHGSSTPFEYLEDIEPPRPILPSIPYPSATNPLYPAAFGGVNMWSSMGQPTDVY 483

Qy        144 ------QQRQLGKKK-----HRRRPSKK-----KRHWKPYYKLTWEEKKKFDEKQSL--- 184
                   :|: : ::      | || :::     |     | :|| : :::  |:: |
Db        484 MMQTNFEQQLVAERNRLLDDHNRRLAEEAEKMAKFQEAMYRQLTMQHEQRVREQELLLQK 543

Qy        185 RASRI-RAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDDFME 243
              ||  | : |  :| :  |        :   :||  :::     ::|| :::|    :
Db        544 RAEEIEKREAESKREEAA-------RLQKLEQEAREIE----ERKAALEAEDRRKREI-- 590

Qy        244 EGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELI-KEYLELEKCLSRME 302
                                    | | |     :    | |:| ||  :  :     |
Db        591 ------------------------EEYNR-----MCEKKKNEIIAKEAADRMRLEEATE 620

Qy        303 DENNRLRLESKRLGGDDARVRELELELDRLRAENLQLLTENELHRQQERAPLSK 356
              |  ||  ||         || | ::  ||:|||       | |  |  ||  :::
Db        621 RERRRLEAES--------RVAEEKIRRDRVRAELEAREREEERKRAAERERIAR 666


Search completed: February  5, 2004, 13:31:59
Job time : 48 secs
```

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: February 5, 2004, 12:05:33 ; Search time 25 Seconds
(without alignments)
675.303 Million cell updates/sec

Title: US-09-972-758A-2
Perfect score: 1910
Sequence: 1 MAEPFLSEYQHQPQTSNCTG.....LTENELHRQQERAPLSKFGD 359

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	164.5	8.6	1224	1	DYNA_CHICK	P35458 gallus gall
2	139	7.3	756	1	BFS1_BOVIN	Q06002 bos taurus
3	137	7.2	536	1	GAG_MLVCB	P27460 cas-br-e mu
4	134.5	7.0	1181	1	NKX1_RAT	Q9qzm6 rattus norv
5	134	7.0	568	1	INVO_RAT	P48998 rattus norv
6	133	7.0	538	1	GAG_MLVMO	P03332 moloney mur
7	133	7.0	1493	1	ERC6_HUMAN	Q03468 homo sapien
8	131	6.9	1898	1	TRHY_HUMAN	Q07283 homo sapien
9	130	6.8	466	1	CMGA_RAT	P10354 rattus norv
10	129.5	6.8	1816	1	AF6_HUMAN	P55196 homo sapien
11	128	6.7	1733	1	VNUA_PRVKA	P33485 pseudorabie
12	128	6.7	2476	1	ATRX_MOUSE	Q61687 mus musculu
13	127.5	6.7	4687	1	PLE1_RAT	P30427 rattus norv
14	126.5	6.6	1407	1	TRHY_RABIT	P37709 oryctolagus
15	125.5	6.6	773	1	DGR8_HUMAN	Q8wyq5 homo sapien
16	125.5	6.6	1549	1	TRHY_SHEEP	P22793 ovis aries
17	125	6.5	2004	1	MOZ_HUMAN	Q92794 homo sapien
18	124.5	6.5	332	1	FA9A_HUMAN	Q8izu1 homo sapien
19	124	6.5	449	1	CMGA_BOVIN	P05059 bos taurus
20	123.5	6.5	677	1	SG1_MOUSE	P16014 mus musculu
21	123.5	6.5	1850	1	BA2A_MOUSE	Q91ye5 mus musculu
22	122.5	6.4	4473	1	PLE1_CRIGR	Q9ji55 cricetulus
23	122	6.4	723	1	SSRP_DROME	Q05344 drosophila
24	121	6.3	538	1	GAG_MSVMO	P03334 moloney mur
25	121	6.3	726	1	BRD3_HUMAN	Q15059 homo sapien
26	121	6.3	1089	1	Y553_HUMAN	Q9ukj3 homo sapien
27	121	6.3	1280	1	DYNA_RAT	P28023 rattus norv
28	120	6.3	536	1	GAG_MLVDE	P29168 murine leuk
29	120	6.3	699	1	NP14_HUMAN	Q14978 homo sapien
30	120	6.3	1781	1	AK12_HUMAN	Q02952 homo sapien
31	119.5	6.3	556	1	HIR3_HUMAN	Q9bw71 homo sapien
32	119.5	6.3	919	1	INCE_HUMAN	Q9nqs7 homo sapien
33	118.5	6.2	472	1	YWIE_CAEEL	Q23525 caenorhabdi
34	118.5	6.2	764	1	UBF1_RAT	P25977 rattus norv
35	117.5	6.2	372	1	TOLA_HAEIN	P44678 haemophilus
36	117.5	6.2	513	1	DMP1_HUMAN	Q13316 homo sapien
37	117.5	6.2	2142	1	BAT2_HUMAN	P48634 homo sapien
38	117.5	6.2	5147	1	PCLO_HUMAN	Q9y6v0 homo sapien
39	117	6.1	568	1	RGS3_MOUSE	Q9dc04 mus musculu
40	117	6.1	845	1	NFM_RAT	P12839 rattus norv
41	117	6.1	911	1	CAFA_MOUSE	Q9qwf0 mus musculu
42	116.5	6.1	503	1	DMP1_MOUSE	O55188 mus musculu
43	116.5	6.1	1281	1	DYNA_MOUSE	O08788 mus musculu
44	116	6.1	927	1	SLA2_CAEEL	Q02328 caenorhabdi
45	116	6.1	1729	1	TABP_HUMAN	Q9c0c2 homo sapien

ALIGNMENTS

RESULT 1
DYNA_CHICK
ID DYNA_CHICK STANDARD; PRT; 1224 AA.
AC P35458;
DT 01-JUN-1994 (Rel. 29, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dynactin 1 (150 kDa dynein-associated polypeptide) (DP-150) (DAP-150)
DE (p150-glued).
GN DCTN1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Cleveland D.W.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE OF 172-1224 FROM N.A.
RC TISSUE=Embryonic brain;
RX MEDLINE=92098576; PubMed=1836789;
RA Gill S.R., Schroer T.A., Szilak I., Steuer E.R., Sheetz M.P.,
RA Cleveland D.W.;
RT "Dynactin, a conserved, ubiquitously expressed component of an
RT activator of vesicle motility mediated by cytoplasmic dynein.";
RL J. Cell Biol. 115:1639-1650(1991).
CC -!- FUNCTION: DYNACTIN IS A MAJOR COMPONENT OF ACTIVATOR I, A 20S
CC POLYPEPTIDE COMPLEX THAT STIMULATES DYNEIN-MEDIATED VESICLE
CC TRANSPORT.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=1;
CC Comment=At least 3 isoforms are produced;
CC Name=1;
CC IsoId=P35458-1; Sequence=Displayed;
CC -!- TISSUE SPECIFICITY: Ubiquitously expressed.
CC -!- SIMILARITY: BELONGS TO THE DYNACTIN 150 KDA SUBUNIT FAMILY.
CC -!- SIMILARITY: Contains 1 CAP-Gly domain.
CC ---

CC ---
DR EMBL; X62773; CAA44617.2; -.
DR PIR; A41642; A41642.
DR InterPro; IPR000938; CAP-Gly.
DR Pfam; PF01302; CAP_GLY; 1.
DR PROSITE; PS00845; CAP_GLY_1; 1.
DR PROSITE; PS50245; CAP_GLY_2; 1.
KW Motor protein; Microtubules; Dynein; Coiled coil; Cytoskeleton;

```
KW   Alternative splicing.
FT   DOMAIN       49     91       CAP-GLY.
FT   DOMAIN      205    540       COILED COIL (POTENTIAL).
FT   DOMAIN      936   1042       COILED COIL (POTENTIAL).
FT   DOMAIN     1081   1117       COILED COIL (POTENTIAL).
SQ   SEQUENCE   1224 AA;  135562 MW;  03B7FFE68E7C01D7 CRC64;

  Query Match             8.6%;  Score 164.5;  DB 1;  Length 1224;
  Best Local Similarity   22.6%;  Pred. No. 0.045;
  Matches   83;  Conservative   61;  Mismatches  152;  Indels   71;  Gaps   14;

Qy         29 NPERPPGAEERVPEEDSRWQSRAFPQLGGRPGPEGEGSLESQPPPLQTQACPESSCLREG 88
              :|| |  |  :||:  ||  ::     | :|    : :   :| | :|     ||    |
Db        106 SPETPESAALKVPKRHSRXAAKGSKLRGAKP----KKTTARRPKPTRTPTSAPSS----G 157

Qy         89 EKGQNGDDSSAGGDF--PPPAEVEPTP-EAELLAQPCHDSEASKLGAPAAGGEEEWGQQQ 145
               | :|  |::||:    |:    ||  | ::  |   |  : :     ||    |
Db        158 TAGPSGSASASGGEMSSSEPSTPAQTPLVAPVIPSPSLTSPVAPMVPSPTKEEENLRSQV 217

Qy        146 RQLGKK----KHRRRPSKKKRHWKPYYKLTWEE----KKKFDEKQSLRASRIR-AEMFAK 196
              | | :|    | :|   | |      ||:  |:    | |  |:|:    |:: |:  ||
Db        218 RDLEEKLETLKIKRNEDKAKLKELEKYKIQLEQVQEWKSKMQEQQADLQRRLKEAKKEAK 277

Qy        197 GQPVAPYNTTQFLMDDHDQ-EEPDLKTGLYSKRAAAKSDDTSDDDFMEEGGE-------- 247
                  |    : : |  |  |   |  :  :|| :   :    | ::|  |
Db        278 DALEAKERYMEEMADTADAIEMATLDKEMAEERAESLQQEV---DSLKEKVEYLTMDLEI 334

Qy        248 -----EDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIK----------EYL 292
                 |: ||||              : :|:    |      |: |::          |::
Db        335 LKHEIEEKGSDGA-----------ASSYQVKQLEEQNARLKEALVRMRDLSASEKQEHV 382

Qy        293 ELEKCLSRMEDENNRLRLESKRLGGDDARVRELELELDRLR---------AENLQLLTEN 343
              :|:| : :   |   || : ::|      |:: |  :| |:          | :: |||
Db        383 KLQKQMEKKNTELESLRQQREKL---QEEVKQAEKTVDELKEQVDAALGAEEMVETLTER 439

Qy        344 ELHRQQE 350
               |  :::
Db        440 NLDLEEK 446


RESULT 2
BFS1_BOVIN
ID   BFS1_BOVIN     STANDARD;      PRT;   756 AA.
AC   Q06002;
DT   01-OCT-1996 (Rel. 34, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Filensin (Beaded filament structural protein 1).
GN   BFSP1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Lens;
RX   MEDLINE=93260017; PubMed=8491777;
RA   Gounari F., Merdes A., Quinlan R., Hess J.F., Fitzgerald P.G.,
RA   Ouzounis C.A., Georgatos S.D.;
RT   "Bovine filensin possesses primary and secondary structure similarity
RT   to intermediate filament proteins.";
RL   J. Cell Biol. 121:847-853(1993).
RN   [2]
RP   REVISIONS, AND SEQUENCE FROM N.A.
RC   TISSUE=Lens;
RA   Hess J.F.;
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: ASSOCIATES WITH BFSP2.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE- AND CYTOSKELETON-ASSOCIATED.
CC   -!- TISSUE SPECIFICITY: LENS.
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X72388; CAA51081.1; -.
DR   InterPro; IPR001664; IF.
DR   PROSITE; PS00226; IF; FALSE_NEG.
KW   Intermediate filament; Repeat; Membrane; Coiled coil; Cytoskeleton;
KW   Eye lens protein; Phosphorylation.
FT   DOMAIN       1     38       HEAD.
FT   DOMAIN      39    318       ROD.
FT   DOMAIN     319    755       TAIL.
FT   DOMAIN      39     73       COIL 1A.
FT   DOMAIN      74     82       LINKER 1.
FT   DOMAIN      83    182       COIL 1B.
FT   DOMAIN     183    199       LINKER 12.
FT   DOMAIN     200    318       COIL 2.
FT   DOMAIN     531    621       7 X 14 AA TANDEM REPEATS.
FT   REPEAT     531    544       1.
FT   REPEAT     545    551       2 (INCOMPLETE).
FT   REPEAT     552    565       3.
FT   REPEAT     566    579       4.
FT   REPEAT     580    593       5.
FT   REPEAT     594    607       6.
FT   REPEAT     608    621       7.
FT   MOD_RES      5      5       PHOSPHORYLATION (BY PKA) (POTENTIAL).
SQ   SEQUENCE   756 AA;  83001 MW;  F86A18208A8E6109 CRC64;

  Query Match             7.3%;  Score 139;  DB 1;  Length 756;
  Best Local Similarity   22.6%;  Pred. No. 0.67;
  Matches   88;  Conservative   32;  Mismatches  117;  Indels  152;  Gaps   16;

Qy         26 EELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPGPEGEGSLESQPP-PLQTQA------ 78
              || :|    || | ||  |     |:|| :||     : :|  | :||  | |:
Db        415 EEASPPTQEGAPEDVP--DGGKISKAFEKLGKMIKEKVKGPKEPEPPADLYTKGRYVMVS 472

Qy         79 ----------CPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTP--------EAELLAQ 120
                       |  |    : |      ||||      || : |||:|        | :   |
Db        473 GDGSFVDPGPCVFSVPAKGGVVVSKGDDS-----VPPDSGVEPSPQQPEPPLEEGQGPPQ 527

Qy        121 PCHDSEASKLGAPAAGGE-----------------------------------------EEWG 142
                  |    : | |   ||                                         || |
Db        528 EKEDGLKEEGGPPEGKGEPPEGKGDSVKEEGGPPEGKGDGVKEEGGPPEGKGDGVKEEGG 587

Qy        143 QQQ-RQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDEKQSLRASRIRAEMFAKGQPVA 201
                : :  | ||    |  |    |       ||:    ||:              ||
Db        588 PPEGKGDGVKKEGEPPEGKGEGLK-------EEEGPLQEKED-------------GQSPT 627

Qy        202 PYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDDFMEEG----------GEED-- 249
              |:       | |::      ||  |:          ||  |||          | |
Db        628 PHPA-----DKGDEKNAKELKGLQGKQ----------DDQKEEGARGPCPMVAPGPEGPS 672

Qy        250 -----------GGSDGMGGDGSEFLQR-----------DFSETYERYHTESLQNMSKQEL 287
                        |||:| |      | |           :  |: |:: |||:|   :  :
Db        673 TPRSQGPQVILGGSEGHGARSGSRLARSPPRKLAYEKVEVMESIEKFSTESIQTYEETAV 732

Qy        288 IKEYLELEKCLSRMEDENNRLRLESKRLG 316
              | | : :||           :   |:||
Db        733 IVETM-IEK-----------TKANKKKLG 749


RESULT 3
GAG_MLVCB
ID   GAG_MLVCB      STANDARD;      PRT;   536 AA.
AC   P27460;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   GAG polyprotein [Contains: Core protein P15; Inner coat protein P12;
```

```
DE   Core shell protein P30; Nucleoprotein P10].
GN   GAG.
OS   Cas-Br-E murine leukemia virus.
OC   Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
OX   NCBI_TaxID=11792;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91227170; PubMed=1840655;
RA   Perryman S.M., McAtee F.J., Portis J.L.;
RT   "Complete nucleotide sequence of the neurotropic murine retrovirus
RT   CAS-BR-E.";
RL   Nucleic Acids Res. 19:1707-1707(1991).
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-POL
CC       POLYPROTEIN.
CC   -!- SIMILARITY: Contains 1 CCHC-type zinc finger.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X57540; CAA40759.1; -.
DR   InterPro; IPR000840; Gag_MA.
DR   InterPro; IPR002079; Gag_p12.
DR   InterPro; IPR003036; Gag_p30.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF01140; Gag_MA; 1.
DR   Pfam; PF01141; Gag_p12; 1.
DR   Pfam; PF02093; Gag_p30; 1.
DR   Pfam; PF00098; zf-CCHC; 1.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   SMART; SM00343; ZnF_C2HC; 1.
DR   PROSITE; PS50158; ZF_CCHC; 1.
KW   Core protein; Coat protein; Nucleoprotein; Polyprotein; Myristate;
KW   Zinc-finger.
FT   CHAIN         2    129       CORE PROTEIN P15.
FT   CHAIN       130    214       INNER COAT PROTEIN P12.
FT   CHAIN       215    477       CORE SHELL PROTEIN P30.
FT   CHAIN       478    536       NUCLEOPROTEIN P10.
FT   ZN_FING     500    517       CCHC-TYPE.
FT   LIPID         2      2       MYRISTATE (BY SIMILARITY).
SQ   SEQUENCE   536 AA;  60740 MW;  3E0F863A393960AE CRC64;

  Query Match             7.2%;  Score 137;  DB 1;  Length 536;
  Best Local Similarity   21.3%;  Pred. No. 0.6;
  Matches   88;  Conservative   54;  Mismatches  141;  Indels  130;  Gaps   19;

Qy         30 PERPPGAEERVPEEDSRWQSRA------FPQLGGRPGPE---GEGS-----LESQPPPLQ 75
              |  || |   :||       |:       | || :| |:     |       |   ||| :
Db        105 PPLPPSAPSLLPEPPLSTSPRSSLYPALTPSLGAKPKPQVLPDSGGPLIDLLTEDPPPYR 164

Qy         76 TQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELL----AQPCHDSEASKLG 131
                | |      :: ::  ::: |: |     :|:| | |      |  ||  |:
Db        165 DPGPPPS------DRDRDDGEAAPAGEAP-----DPSPMASRLRGRRELPVADSTTSQAF 213

Qy        132 APAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWK---PYY-----KL------------- 170
                 :||           |: ::    |    :||   | :     ||
Db        214 PLRSGGN----------GQLQYWPFSSSDLYNWKNNNPSFSEDPGKLTALIESVLLTHQP 263

Qy        171 TWEEKKKF-------DEKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPD---- 219
              ||:: ::        :||| :    : |    :|:   |       : |    | ||
Db        264 TWDDCQQLLGTLLTGEEKQRV---LLEARKAVRGEDGRPTQLPNEINDAFPLERPDWDYN 320

Qy        220 --------------LKTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQR 265
                           |  || :   | :|             :  | : |     | ||:|
Db        321 TQRGRNHLVLYRQLLLAGLQN---AGRSPTNL--------AKVKGITQGPNESPSAFLER 369

Qy        266 DFSETYERYHTESLQ------NMSKQELIKEYLELEKCLSRMEDENNRLRLESKRLGG-- 317
                 | | ||     :      |:|   : :   :: : | |:||      |:|| ||
```

```
Db        370 -LKEAYRRYTPYDPEDPGQETNVSMSFIWQSAPDIGRKLERLED------LKSKTLGDLV 422

Qy        318 ---------------DDARVRELELELDRLRAENLQLLTENELHRQQERAPL 354
                            ::   || | : :| |||: |   | :  | :| : |
Db        423 REAEKIFNKRETPEEREERIKRETEEKEERRRAEDEQKEKERDRRRHREMSKL 475


RESULT 4
NKX1_RAT
ID   NKX1_RAT       STANDARD;      PRT;  1181 AA.
AC   Q9QZM6; Q62932;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Sodium/potassium/calcium exchanger 1 precursor (Na(+)/K(+)/Ca(2+)-
DE   exchange protein 1) (Retinal rod Na-Ca+K exchanger).
GN   SLC24A1 OR NCKX1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   STRAIN=Sprague-Dawley; TISSUE=Eye;
RX   MEDLINE=20217335; PubMed=10751314;
RA   Poon S., Leach S., Li X.-F., Tucker J.E., Schnetkamp P.P.M.,
RA   Lytton J.;
RT   "Alternatively spliced isoforms of the rat eye
RT   sodium/calcium+potassium exchanger NCKX1.";
RL   Am. J. Physiol. 278:C651-C660(2000).
RN   [2]
RP   SEQUENCE OF 1067-1155 FROM N.A.
RA   White K.E., Gesek F.A., Friedman P.A.;
RL   Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Critical component of the visual transduction cascade,
CC       controlling the calcium concentration of outer segments during
CC       light and darkness. Light causes a rapid lowering of cytosolic
CC       free calcium in the outer segment of both retinal rod and cone
CC       photoreceptors and the light-induced lowering of calcium is caused
CC       by extrusion via this protein which plays a key role in the
CC       process of light adaptation. Transports one Ca(2+) and one K(+) in
CC       exchange for four Na(+).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=4;
CC       Name=1;
CC         IsoId=Q9QZM6-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q9QZM6-2; Sequence=VSP_006161;
CC       Name=3;
CC         IsoId=Q9QZM6-3; Sequence=VSP_006162;
CC       Name=4;
CC         IsoId=Q9QZM6-4; Sequence=VSP_006163;
CC   -!- TISSUE SPECIFICITY: Highly expressed in the eye.
CC   -!- SIMILARITY: BELONGS TO THE SLC24A FAMILY OF TRANSPORTERS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF176688; AAD53121.1; -.
DR   EMBL; U49235; AAB37753.1; -.
DR   InterPro; IPR004817; K_NaCaexchang.
DR   InterPro; IPR004481; K_NaCaexchng.
DR   InterPro; IPR004837; NaCa_Exmemb.
DR   Pfam; PF01699; Na_Ca_Ex; 2.
DR   TIGRFAMs; TIGR00927; 2A1904; 1.
DR   TIGRFAMs; TIGR00367; TIGR00367; 1.
KW   Vision; Transport; Antiport; Symport; Calcium transport;
```

```
KW   Transmembrane; Glycoprotein; Phosphorylation; Signal; Repeat;
KW   Alternative splicing.
FT   SIGNAL        1     38       POTENTIAL.
FT   CHAIN        39   1181       SODIUM/POTASSIUM/CALCIUM EXCHANGER 1.
FT   DOMAIN       39    419       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    420    440       POTENTIAL.
FT   DOMAIN      441    464       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    465    485       POTENTIAL.
FT   DOMAIN      486    491       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    492    512       POTENTIAL.
FT   DOMAIN      513    519       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    520    544       POTENTIAL.
FT   DOMAIN      545    552       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    553    569       POTENTIAL.
FT   DOMAIN      570    989       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    990   1010       POTENTIAL.
FT   DOMAIN     1011   1017       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1018   1038       POTENTIAL.
FT   DOMAIN     1039   1053       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1054   1074       POTENTIAL.
FT   DOMAIN     1075   1092       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1093   1113       POTENTIAL.
FT   DOMAIN     1114   1121       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1122   1142       POTENTIAL.
FT   DOMAIN     1143   1150       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1151   1171       POTENTIAL.
FT   DOMAIN     1172   1181       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      461    501       ALPHA-1.
FT   REPEAT     1061   1092       ALPHA-2.
FT   DOMAIN      730    905       14 X APPROXIMATE TANDEM REPEATS.
FT   REPEAT      730    741       1.
FT   REPEAT      742    754       2.
FT   REPEAT      755    766       3.
FT   REPEAT      767    778       4.
FT   REPEAT      779    791       5.
FT   REPEAT      792    804       6.
FT   REPEAT      805    817       7.
FT   REPEAT      818    830       8.
FT   REPEAT      831    843       9.
FT   REPEAT      844    856       10.
FT   REPEAT      857    869       11.
FT   REPEAT      870    881       12.
FT   REPEAT      882    893       13.
FT   REPEAT      894    905       14.
FT   DOMAIN      952    974       POLY-GLU.
FT   MOD_RES     625    625       PHOSPHORYLATION (POTENTIAL).
FT   CARBOHYD    271    271       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    598    710       Missing (in isoform 2).
FT                                /FTId=VSP_006161.
FT   VARSPLIC    616    710       Missing (in isoform 3).
FT                                /FTId=VSP_006162.
FT   VARSPLIC    652    679       Missing (in isoform 4).
FT                                /FTId=VSP_006163.
SQ   SEQUENCE   1181 AA;  129980 MW;  B063C1C1193696AE CRC64;

  Query Match             7.0%;  Score 134.5;  DB 1;  Length 1181;
  Best Local Similarity   22.5%;  Pred. No. 1.9;
  Matches   67;  Conservative   44;  Mismatches  160;  Indels   27;  Gaps    7;

Qy         19 TGAAAVQEELNPERPPGAEERVPEEDSRW-----QSRAFPQLGGRPGPEGEGSLESQPPP 73
              | | | ::: : |  || :| | | : |      :    :  |:   |||   | :
Db        692 TPAPAPEDKGDQEEDPGCQEDVDEAEHRGDMTGEEGERETEAEGKKDEEGETEAERKEDG 751

Qy         74 LQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELLAQPC-HDSEASKLG- 131
               : :   :    :|||    | |   |       | :   | |   :   |: |    |
Db        752 QEEETETKGKEKQEGETESEGKDEQEGETEAEGKEADHEGETEAEGKEVEHEGETEAEGT 811

Qy        132 -------APAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDEKQSL 184
                    | | | |   :    ||:      :: :|    :   |  | |: | :
Db        812 EDEQEGETEAEGKEVEQEGETEAEGKEVEHEVETEAERKETNHEGETEAEGKEADHEGET 871

Qy        185 RAS---RIRAEMFAKGQPVAPYNTTQFLMDDHD------QEEPDLKTGLYSKRAAAKSDD 235
                  |      : |  |:|:       |     |:|:       :: ::| |     | | ::|
```

```
Db        872 EAEGNVEHQGETEAEGKVEHEGETEAGEKDEHEGQSETQADDTEVKDG--EGEAEANAED 929

Qy        236 TSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLE 293
                  :    |:| :  |||| |||  |    :  | |    |  :  |:: |  | : |
Db        930 QCETAQGEKGADGGGGSD--GGDSEEEEDEEDEEEEEEEEEEEEEESEEPLSLEWPE 985


RESULT 5
INVO_RAT
ID   INVO_RAT       STANDARD;      PRT;   568 AA.
AC   P48998;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Involucrin.
GN   IVL.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley;
RX   MEDLINE=94104476; PubMed=8277848;
RA   Djian P., Phillips M., Easley K., Huang E., Simon M., Rice R.H.,
RA   Green H.;
RT   "The involucrin genes of the mouse and the rat: study of their shared
RT   repeats.";
RL   Mol. Biol. Evol. 10:1136-1149(1993).
CC   -!- FUNCTION: Part of the insoluble cornified cell envelope (CE) of
CC       stratified squamous epithelia.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic. Constituent of the scaffolding
CC       of the cornified envelope.
CC   -!- TISSUE SPECIFICITY: Keratinocytes of epidermis and other
CC       stratified squamous epithelia.
CC   -!- PTM: Substrate of transglutaminase. Specific glutamines or lysines
CC       are cross-linked to keratins, desmoplakin and to inter involucrin
CC       molecules.
CC   -!- SIMILARITY: BELONGS TO THE INVOLUCRIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L28818; AAA41445.1; -.
DR   PIR; I61106; I61106.
DR   InterPro; IPR002360; Involucrin.
DR   PROSITE; PS00795; INVOLUCRIN; 1.
KW   Keratinization; Repeat.
SQ   SEQUENCE   568 AA;  67021 MW;  E2B07740FE779345 CRC64;

  Query Match             7.0%;  Score 134;  DB 1;  Length 568;
  Best Local Similarity   20.9%;  Pred. No. 0.94;
  Matches   93;  Conservative   66;  Mismatches  169;  Indels  118;  Gaps   18;

Qy          3 EPFLSEYQHQPQTSNCTGAAAVQEE-----------------LNPERPPGAEERVPEED 44
              | | ::|  | |    |    |::                   || | | ::: |||
Db        127 EQHLRQHQQPQQESQGQGLCLGQQQDVLAPQELHMGQHQKEKLQEPELPLGQQQKTPEEQ 186

Qy         45 -----SRWQSRAFPQLGGRPGPE----GEGSLESQPPPLQTQAC---------PESSCLR 86
                   : |      :    | :    |:  : ||   : |              || |||
Db        187 ELILGEKQQKLHLVERHQEPQEQELHHGQKQKQQQPQEQELQLVQHQKQKQHEPE-LCLR 245

Qy         87 EGEKGQNGDDSSAGGDFPPPAEVEP------------TPEAELLAQPCHDSEASKLGAP 133
              : :: :: :      |          ||               ||  |  |   :|   :|
Db        246 KQQQQESHERELHLGKQQQQESHEPELHLGKQQHQESHEPELHLGKQQHQESCEPEL--- 302

Qy        134 AAGGEEEWGQQQRQ--------LGKKKHRR------RPSKKKRHWKPYYKLTWEEKKKF- 178
                | ::   ||| |        |||::|:       :  |:::  :|   |  :||:|
```

```
Db        303 HLGEQQHQEQQQHQESCEPELHLGKQQHQETQESELQLGKQQKPHEPDMVLDPKEKQKLH 362

Qy        179 --------DEKQSLRASRIRAEM------------FAKGQPVAPYNTTQFLMDDHDQEEP 218
                      :  |  : | :: |                | | :   : |:   :     ||
Db        363 DPELHLGKQQHQESQESELQVEKKQHEKSPEPELHLGKQQELHEPDMTEDQKEKQSLHEP 422

Qy        219 DLKTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTES 278
              :|  |         |  :: : |  |:   |:     :       | :  |:    | |:|  ::|
Db        423 ELHLG--------KQQESHEPDMTEDQKEKQSLYEPELHLGKQQEQQIEYEGYQR--SKS 472

Qy        279 LQNMSKQELIKEYLELEKCLSRMEDENNRL--RLESKRLGGDDARVRELELELDRLRAEN 336
              |  : |||       ||:   | :| |   |  ||: : :  |:      :||  : : ||
Db        473 LNQLLKQEKASRGQELDD--SHLEQEKELLDQRLDQELVNKDE------QLERKKHKLEN 524

Qy        337 L--------QLLTENELHRQQERAPL 354
              |        ||:   :  | ||  |:
Db        525 LTQKEKQIKQLVPSTD--RVQETQPI 548


RESULT 6
GAG_MLVMO
ID   GAG_MLVMO      STANDARD;      PRT;   538 AA.
AC   P03332;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   GAG polyprotein [Contains: Core protein P15; Inner coat protein P12;
DE   Core shell protein P30; Nucleoprotein P10].
GN   GAG.
OS   Moloney murine leukemia virus.
OC   Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
OX   NCBI_TaxID=11801;
RN   [1]
RP   SEQUENCE FROM N.A. (CLONE PMLV-1).
RX   MEDLINE=82035843; PubMed=6169994;
RA   Shinnick T.M., Lerner R.A., Sutcliffe J.G.;
RT   "Nucleotide sequence of Moloney murine leukaemia virus.";
RL   Nature 293:543-548(1981).
RN   [2]
RP   SEQUENCE OF 2-31, AND MYRISTOYLATION.
RX   MEDLINE=83169654; PubMed=6340098;
RA   Henderson L.E., Krutzsch H.C., Oroszlan S.;
RT   "Myristyl amino-terminal acylation of murine retrovirus proteins: an
RT   unusual post-translational proteins modification.";
RL   Proc. Natl. Acad. Sci. U.S.A. 80:339-343(1983).
RN   [3]
RP   SEQUENCE OF 479-529.
RX   MEDLINE=81264245; PubMed=6267042;
RA   Henderson L.E., Copeland T.D., Sowder R.C., Smythers G.W.,
RA   Oroszlan S.;
RT   "Primary structure of the low molecular weight nucleic acid-binding
RT   proteins of murine leukemia viruses.";
RL   J. Biol. Chem. 256:8400-8406(1981).
CC   -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-POL
CC       POLYPROTEIN.
CC   -!- SIMILARITY: Contains 1 CCHC-type zinc finger.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J02255; AAB59942.1; -.
DR   PIR; A03930; FOMV1M.
DR   PDB; 1A6B; 02-NOV-99.
DR   PDB; 1MN8; 14-JAN-03.
DR   InterPro; IPR000840; Gag_MA.
DR   InterPro; IPR002079; Gag_p12.
DR   InterPro; IPR003036; Gag_p30.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF01140; Gag_MA; 1.
DR   Pfam; PF01141; Gag_p12; 1.
DR   Pfam; PF02093; Gag_p30; 1.
DR   Pfam; PF00098; zf-CCHC; 1.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   SMART; SM00343; ZnF_C2HC; 1.
DR   PROSITE; PS50158; ZF_CCHC; 1.
KW   Core protein; Coat protein; Nucleoprotein; Polyprotein; Myristate;
KW   Zinc-finger; 3D-structure.
FT   CHAIN         2    131       CORE PROTEIN P15.
FT   CHAIN       132    215       INNER COAT PROTEIN P12.
FT   CHAIN       216    478       CORE SHELL PROTEIN P30.
FT   CHAIN       479    534       NUCLEOPROTEIN P10.
FT   ZN_FING     502    519       CCHC-TYPE.
FT   LIPID         2      2       MYRISTATE.
SQ   SEQUENCE   538 AA;  60858 MW;  8A7652439B464495 CRC64;

  Query Match             7.0%;  Score 133;  DB 1;  Length 538;
  Best Local Similarity   20.9%;  Pred. No. 1;
  Matches   89;  Conservative   40;  Mismatches  142;  Indels  154;  Gaps   19;

Qy         30 PERPPGA-------EERVPEEDSRWQSRAFPQLGGRPGPE---GEGS-----LESQPPPL 74
              |  || |       |    | : :   | || :| |:     |      |   |||
Db        107 PPLPPSAPSLPLEPPRSTPPRSSLYPALT-PSLGAKPKPQVLSDSGGPLIDLLTEDPPPY 165

Qy         75 QTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELL----AQPCHDSEASKL 130
              :    | |   |:|    || :::  |: |     :|:| | |       |  ||  |:
Db        166 RDPRPPPSD--RDG----NGGEATPAGEAP-----DPSPMASRLRGRREPPVADSTTSQA 214

Qy        131 GAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDEKQSLRASRIR 190
                 |||              |: ::    |    :||             | |      : |
Db        215 FPLRAGGN----------GQLQYWPFSSSDLYNWK-------NNNPSFSEDPGKLTALIE 257

Qy        191 AEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTG------LYSKRAAAKSDD--------- 235
              : :           | |   ||  |   | ||       |  | | : ||
Db        258 SVLI----------THQPTWDDCQQLLGTLLTGEEKQRVLLEARKAVRGDDGRPTQLPNE 307

Qy        236 --------TSDDDFMEEGGEE-------------------------DGGSDGMGGDGSE 261
                      | |:  : |                           | : |     |
Db        308 VDAAFPLERPDWDYTTQAGRNHLVHYRQLLLAGLQNAGRSPTNLAKVKGITQGPNESPSA 367

Qy        262 FLQRDFSETYERY-----------------------------HTESLQNMSKQELIKE 290
              ||:|   | | ||                             | |:| :  :|::
Db        368 FLER-LKEAYRRYTPYDPEDPGQETNVSMSFIWQSAPDIGRKLERLEDLKNKTLGDLVR- 425

Qy        291 YLELEKCLSRMEDENNRLRLESKRLGGDDARV-RELELELDRLRAENLQLLTENELHRQQ 349
              | ||  :: |    |           : |: || | : :| | |:|  |  : |  | :
Db        426 --EAEKIFNKRETPEER----------EERIRRETEEKEERRRTEDEQKEKERDRRRHR 472

Qy        350 ERAPL 354
              | : |
Db        473 EMSKL 477


RESULT 7
ERC6_HUMAN
ID   ERC6_HUMAN     STANDARD;      PRT;  1493 AA.
AC   Q03468;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Excision repair protein ERCC-6 (Cockayne syndrome protein CSB).
GN   ERCC6 OR CSB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93092214; PubMed=1339317;
RA   Troelstra C., van Gool A., de Wit J., Vermeulen W., Bootsma D.,
RA   Hoeijmakers J.H.J.;
RT   "ERCC6, a member of a subfamily of putative helicases, is involved in
```

```
RT   Cockayne's syndrome and preferential repair of active genes.";
RL   Cell 71:939-953(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93181229; PubMed=8382798;
RA   Troelstra C., Hesen V., Bootsma D., Hoeijmakers J.H.J.;
RT   "Structure and expression of the excision repair gene ERCC6, involved
RT   in the human disorder Cockayne's syndrome group B.";
RL   Nucleic Acids Res. 21:419-426(1993).
RN   [3]
RP   REVIEW ON VARIANTS CSB.
RX   MEDLINE=99374920; PubMed=10447254;
RA   Cleaver J.E., Thompson L.H., Richardson A.S., States J.C.;
RT   "A summary of mutations in the UV-sensitive disorders: Xeroderma
RT   pigmentosum, Cockayne syndrome, and trichothiodystrophy.";
RL   Hum. Mutat. 14:9-22(1999).
RN   [4]
RP   VARIANTS CSB, AND VARIANTS.
RX   MEDLINE=98107940; PubMed=9443879;
RA   Mallery D.L., Tanganelli B., Colella S., Steingrimsdottir H.,
RA   van Gool A.J., Troelstra C., Stefanini M., Lehmann A.R.;
RT   "Molecular analysis of mutations in the CSB (ERCC6) gene in patients
RT   with Cockayne syndrome.";
RL   Am. J. Hum. Genet. 62:77-85(1998).
RN   [5]
RP   DISEASE.
RX   MEDLINE=20206317; PubMed=10739753;
RA   Meira L.B., Graham J.M. Jr., Greenberg C.R., Busch D.B.,
RA   Doughty A.T.B., Ziffer D.W., Coleman D.M., Savre-Train I.,
RA   Friedberg E.C.;
RT   "Manitoba aboriginal kindred with original cerebro-oculo-facio-
RT   skeletal syndrome has a mutation in the Cockayne syndrome group B
RT   (CSB) gene.";
RL   Am. J. Hum. Genet. 66:1221-1228(2000).
RN   [6]
RP   DISEASE.
RX   MEDLINE=20231544; PubMed=10767341;
RA   Colella S., Nardo T., Botta E., Lehmann A.R., Stefanini M.;
RT   "Identical mutations in the CSB gene associated with either Cockayne
RT   syndrome or the DeSanctis-Cacchione variant of xeroderma
RT   pigmentosum.";
RL   Hum. Mol. Genet. 9:1171-1175(2000).
CC   -!- FUNCTION: IS INVOLVED IN THE PREFERENTIAL REPAIR OF ACTIVE GENES.
CC       PRESUMED DNA OR RNA UNWINDING FUNCTION. CORRECTS THE UV SURVIVAL
CC       AND RNA SYNTHESIS AFTER UV EXPOSURE OF COCKAYNE'S SYNDROME
CC       COMPLEMENTATION GROUP B.
CC   -!- SUBUNIT: INTERACTS WITH THE CSB PROTEIN AND A SUBUNIT OF RNA
CC       POLYMERASE II TFIIH.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- DISEASE: Defects in ERCC6 are the cause of Cockayne's syndrome
CC       type B (CSB) [MIM:133540]. CSB is a disease which is characterized
CC       by dwarfism, precociously senile appearance, pigmentary retinal
CC       degeneration, optic atrophy, deafness, sensitivity to sunlight,
CC       and mental retardation.
CC   -!- DISEASE: Defects in ERCC6 are a cause of cerebro-oculo-facio-
CC       skeletal syndrome (COFS) [MIM:214150], a rapidly progressive
CC       neurological disorder leading to brain atrophy with calcification,
CC       cataracts, microcornea, optic atrophy, progressive joint
CC       contractures, and growth failure. Inheritance is autosomal
CC       recessive.
CC   -!- DISEASE: Defects in ERCC6 are a cause of De Sanctis-Cacchione
CC       syndrome (DSC) [MIM:278800]; also known as xerodermic idiocy. DSC
CC       is an autosomal recessive syndrome consisting of xeroderma
CC       pigmentosum associated with mental retardation, retarded growth,
CC       gonadal hypoplasia, and sometimes neurologic complications.
CC   -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC   -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC       WWW="http://www.infobiogen.fr/services/chromcancer/Genes/CSBID302.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L04791; AAA52397.1; -.
DR   PIR; A44224; A44224.
DR   TRANSFAC; T04550; -.
DR   Genew; HGNC:3438; ERCC6.
DR   MIM; 133540; -.
DR   MIM; 214150; -.
DR   MIM; 278800; -.
DR   GO; GO:0003678; F:DNA helicase activity; TAS.
DR   GO; GO:0003702; F:RNA polymerase II transcription factor acti. . .; TAS.
DR   GO; GO:0006281; P:DNA repair; TAS.
DR   GO; GO:0006366; P:transcription from Pol II promoter; TAS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
KW   Helicase; DNA repair; ATP-binding; DNA-binding; Nuclear protein;
KW   Transcription regulation; Disease mutation; Polymorphism;
KW   Cockayne's syndrome; Deafness; Dwarfism.
FT   DOMAIN      356    394       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      442    446       GLY-RICH.
FT   DOMAIN      466    481       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN     1038   1055       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   NP_BIND     532    539       ATP (POTENTIAL).
FT   SITE        646    649       DEGH BOX.
FT   VARIANT     255    255       K -> T.
FT                                /FTId=VAR_001216.
FT   VARIANT     399    399       G -> D.
FT                                /FTId=VAR_001217.
FT   VARIANT     670    670       R -> W (in CSB).
FT                                /FTId=VAR_001218.
FT   VARIANT     851    851       W -> R (in CSB).
FT                                /FTId=VAR_001219.
FT   VARIANT     957    957       V -> G (in CSB).
FT                                /FTId=VAR_001220.
FT   VARIANT    1042   1042       P -> L (in CSB).
FT                                /FTId=VAR_001221.
FT   VARIANT    1095   1095       P -> R (in CSB).
FT                                /FTId=VAR_001222.
FT   VARIANT    1097   1097       M -> V.
FT                                /FTId=VAR_001223.
FT   VARIANT    1213   1213       R -> G (in CSB).
FT                                /FTId=VAR_001224.
FT   VARIANT    1413   1413       Q -> R.
FT                                /FTId=VAR_001225.
SQ   SEQUENCE   1493 AA;  168415 MW;  285257E2AEC071AC CRC64;

  Query Match             7.0%;  Score 133;  DB 1;  Length 1493;
  Best Local Similarity   23.7%;  Pred. No. 2.9;
  Matches   82;  Conservative   42;  Mismatches  118;  Indels  104;  Gaps   17;

Qy         35 GAEERVP------EEDSRWQSRAFPQLGGRPGPEGEGSLESQPPPLQTQACPESSCLREG 88
              ||| ::       |||:             |||    || |   |:|  |  |   :| |
Db        201 GAEVKIELDHASLEEDA-----------EPGP---SSLGSMLMPVQETAWEE--LIRTG 243

Qy         89 EKGQNGDDSSAGGDFPPPAEVEPTP----EA---------------ELLAQPCHDSEASK 129
              :      : |   |   | :|      ||               |   | |:   | |
Db        244 QM------TPFGTQIPQKQEKKPRKIMLNEASGFEKYLADQAKLSFERKKQGCNKRAARK 297

Qy        130 LGAPA-----AGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDEKQSL 184
               ||         : :  :: | | ||: |       |:| |   |   : : |    ::
Db        298 APAPVTPPAPVQNKNKPNKKARVLSKKEER-----LKKHIKKLQKRALQFQGKVGLPKAR 352

Qy        185 RA----SRIRAEMFAKGQPVAPYNTTQFLMDDHDQE----EPDL------------KTG 223
              |      |  ||  ::|:  : |  |:   :: | |    | ||              | |
Db        353 RPWESDMRPEAEGDSEGEE-SEYFPTEEEEEEEDDEVEGAEADLSGDGTDYELKPLPKGG 411
```

```
Qy        224 LYSKRAAAKSDDTSDDDFMEEGGEE-DGGSDGMGGDGSEFLQRDFSETYERYHTESLQNM 282
              |:   :     ||||    ||| :  | | || |   : |   :  | |: : |:
Db        412 KRQKKVPVQE---IDDDFFPSSGEEAEAASVGEGGGGGRKVGRYRDDGDEDYYKQRLRRW 468

Qy        283 SKQELIKEYLELEKCLSRMEDENNRLRLESKRLGGDDARVRELELE 328
              :|               |::|:  ||:||      ||:   : | :
Db        469 NK-------------LRLQDKEKRLKLE------DDSEESDAEFD 494


RESULT 8
TRHY_HUMAN
ID   TRHY_HUMAN     STANDARD;      PRT;  1898 AA.
AC   Q07283;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Trichohyalin.
GN   THH OR TRHY OR THL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93280194; PubMed=7685034;
RA   Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,
RA   Steinert P.M.;
RT   "The structure of human trichohyalin. Potential multiple roles as a
RT   functional EF-hand-like calcium-binding protein, a cornified cell
RT   envelope precursor, and an intermediate filament-associated (cross-
RT   linking) protein.";
RL   J. Biol. Chem. 268:12164-12176(1993).
RN   [2]
RP   SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=93315897; PubMed=7686953;
RA   O'Keefe E.J., Hamilton E.H., Lee S.-C., Steinert P.M.;
RT   "Trichohyalin: a structural protein of hair, tongue, nail, and
RT   epidermis.";
RL   J. Invest. Dermatol. 101:65S-71S(1993).
CC   -!- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
CC       IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
CC       INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
CC       LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
CC       ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
CC       WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
CC       ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
CC       ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
CC       DIFFERENTIATION.
CC   -!- SUBUNIT: Monomer (Probable).
CC   -!- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
CC       THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
CC       THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
CC       THE EPIDERMIS.
CC   -!- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CC       CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
CC       ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
CC       OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED
CC       ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
CC       THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
CC       DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
CC       THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
CC       DIFFERENT SPECIES.
CC   -!- PTM: SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE
CC       PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100
CC       FAMILY.
CC   -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L09190; AAA65582.1; -.
DR   PIR; A45973; A45973.
DR   HSSP; P02633; 4ICB.
DR   Genew; HGNC:11791; THH.
DR   MIM; 190370; -.
DR   GO; GO:0005856; C:cytoskeleton; NAS.
DR   GO; GO:0005509; F:calcium ion binding activity; TAS.
DR   InterPro; IPR001751; CaBP_S100.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR002017; Spectrin.
DR   Pfam; PF00036; efhand; 1.
DR   Pfam; PF01023; S_100; 1.
DR   ProDom; PD003407; CaBP_S100; 1.
DR   PROSITE; PS00018; EF_HAND; 1.
DR   PROSITE; PS00303; S100_CABP; 1.
KW   Keratinization; Calcium-binding; Repeat; Citrullination.
FT   DOMAIN        1     91       S-100 LIKE.
FT   CA_BIND      22     33       EF-HAND 1 (LOW AFFINITY) (POTENTIAL).
FT   CA_BIND      62     73       EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).
FT   DOMAIN      314    390       6 X 13 AA TANDEM REPEATS OF
FT                                R-R-E-Q-E-E-E-R-R-E-Q-Q-L.
FT   REPEAT      314    326       1-1 (APPROXIMATE).
FT   REPEAT      327    339       1-2 (APPROXIMATE).
FT   REPEAT      340    351       1-3 (APPROXIMATE).
FT   REPEAT      352    364       1-4.
FT   REPEAT      365    377       1-5.
FT   REPEAT      378    390       1-6.
FT   DOMAIN      391    444       9 X 6 AA TANDEM REPEATS OF R-R-E-Q-Q-L.
FT   REPEAT      391    396       2-1.
FT   REPEAT      397    402       2-2.
FT   REPEAT      403    408       2-3.
FT   REPEAT      409    414       2-4.
FT   REPEAT      415    420       2-5.
FT   REPEAT      421    426       2-6.
FT   REPEAT      427    432       2-7.
FT   REPEAT      433    438       2-8.
FT   REPEAT      439    444       2-9.
FT   DOMAIN      444    702       9 X 28 AA APPROXIMATE TANDEM REPEATS.
FT   DOMAIN      923   1162       8 X 30 AA TANDEM REPEATS.
FT   REPEAT      923    952       4-1.
FT   REPEAT      953    982       4-2.
FT   REPEAT      983   1012       4-3.
FT   REPEAT     1013   1042       4-4.
FT   REPEAT     1043   1072       4-5.
FT   REPEAT     1073   1102       4-6.
FT   REPEAT     1103   1132       4-7.
FT   REPEAT     1133   1162       4-8.
FT   DOMAIN     1250   1849       23 X 26 AA APPROXIMATE TANDEM REPEATS.
FT   CONFLICT   1752   1752       F -> L (IN REF. 2).
FT   CONFLICT   1794   1801       QERDRQYR -> RSETGSTG (IN REF. 2).
FT   CONFLICT   1857   1857       Q -> K (IN REF. 2).
FT   CONFLICT   1880   1880       V -> G (IN REF. 2).
SQ   SEQUENCE   1898 AA;  247219 MW;  A74B5947FB62E31D CRC64;

  Query Match             6.9%;  Score 131;  DB 1;  Length 1898;
  Best Local Similarity   22.4%;  Pred. No. 4.7;
  Matches   79;  Conservative   54;  Mismatches  146;  Indels   74;  Gaps   13;

Qy         31 ERPPGAEER-------VPEEDSRWQSRAFPQLGGRPGPEGEGSLESQPPPLQTQACPESS 83
              ||   ||||       :|||: : |       ||   | |  |:      | |    :
Db        813 ERQLRAEERQQREQRFLPEEEEKEQR-------GRQRREREKELQFLEEEEQLQRRERAQ 865

Qy         84 CLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAE-------LLAQPCHDSEASKLGAPAAG 136
              |:| | |   |       :    :      | |       | |:|    :  |
Db        866 QLQEEEDGLQEDQERRRQEQRRDQKWRWQLEEERKRRRHTLYAKPALQEQLRKEQQLLQE 925

Qy        137 GEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDEKQSLRASRIRAEMFAK 196
              |||  :::|:      :||  :::| ::   :|  ||::   |::  |  : |   : |
Db        926 EEEELQREERE------KRRRQEQERQYREEEQLQQEEEQLLREEREKRRRQERERQYRK 979
```

```
Qy        197 GQPVAPYNTTQFLMDDHDQ---EEPDLKTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSD 253
Db        980 DKK---------LQQKEEQLLGEEPEKR-----RRQEREKKYREEEELQQE--------- 1016

Qy        254 GMGGDGSEFLQRDFSETYER------YHTESLQNMSKQELIKEYLELEKCLSRMED--EN 305
Db       1017 ------EEQLLREEREKRRRQEWERQYRKKDELQQEEEQLLREEREKRRLQERERQYREE 1070

Qy        306 NRLRLESKRLGGDDARVR---ELELEL---DRLRAENLQLLTEN-ELHRQQER 351
Db       1071 EELQQEEEQLLGEERETRRRQELERQYRKEEELQQEEEQLLREEPEKRRRQER 1123


RESULT 9
CMGA_RAT
ID   CMGA_RAT       STANDARD;      PRT;   466 AA.
AC   P10354;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Chromogranin A precursor (CgA) [Contains: Pancreastatin; Beta-granin;
DE   WE-14].
GN   CHGA.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88112232; PubMed=2828116;
RA   Iacangelo A., Okayama H., Eiden L.E.;
RT   "Primary structure of rat chromogranin A and distribution of its
RT   mRNA.";
RL   FEBS Lett. 227:115-121(1988).
RN   [2]
RP   SEQUENCE OF 29-466 FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=88312980; PubMed=3044825;
RA   Hutton J.C., Nielsen E., Kastern W.;
RT   "The molecular cloning of the chromogranin A-like precursor of beta-
RT   granin and pancreastatin from the endocrine pancreas.";
RL   FEBS Lett. 236:269-274(1988).
RN   [3]
RP   SEQUENCE OF 19-32.
RX   MEDLINE=85285598; PubMed=3896848;
RA   Hutton J.C., Hansen F., Peshavaria M.;
RT   "Beta-granins: 21 kDa co-secreted peptides of the insulin granule
RT   closely related to adrenal medullary chromogranin A.";
RL   FEBS Lett. 188:336-340(1985).
CC   -!- FUNCTION: PANCREASTATIN STRONGLY INHIBITS GLUCOSE INDUCED INSULIN
CC       RELEASE FROM THE PANCREAS.
CC   -!- SUBCELLULAR LOCATION: Neuroendocrine and endocrine secretory
CC       granules.
CC   -!- PTM: CGA IS O-GLYCOSYLATED.
CC   -!- MISCELLANEOUS: BINDS CALCIUM WITH A LOW-AFFINITY.
CC   -!- SIMILARITY: BELONGS TO THE CHROMOGRANIN / SECRETOGRANIN PROTEIN
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X06832; CAA29988.1; -.
DR   HSSP; P05059; 1CFK.
DR   InterPro; IPR001819; Chromogranin_AB.
DR   InterPro; IPR001990; Granin.
DR   Pfam; PF01271; Granin; 1.
DR   PRINTS; PR00659; CHROMOGRANIN.
DR   PROSITE; PS00422; GRANINS_1; 1.
DR   PROSITE; PS00423; GRANINS_2; 1.
KW   Signal; Amidation; Glycoprotein; Calcium-binding.
FT   SIGNAL       1     18
FT   CHAIN       19    466       CHROMOGRANIN A.
FT   PEPTIDE     19    146       BETA-GRANIN.
FT   PEPTIDE    281    332       PANCREASTATIN (POTENTIAL).
FT   PEPTIDE    361    374       WE-14.
FT   DOMAIN      92    110       POLY-GLN.
FT   DOMAIN     231    249       POLY-GLU.
FT   DOMAIN     344    350       POLY-GLU.
FT   MOD_RES    332    332       AMIDATION (G-333 PROVIDE AMIDE GROUP)
FT                               (PROBABLE).
FT   DISULFID    35     56       BY SIMILARITY.
SQ   SEQUENCE   466 AA;  52024 MW;  05D135FFA657C48C CRC64;

  Query Match             6.8%;  Score 130;  DB 1;  Length 466;
  Best Local Similarity   24.0%;  Pred. No. 1.3;
  Matches   87;  Conservative   55;  Mismatches  159;  Indels   62;  Gaps   17;

Qy          3 EPFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGR---P 59
Db         64 ERVLSILRHQ----------NLLKELQDLALQGAKERAQQQQQQQQQQQQQQQQQQQQHS 113

Qy         60 GPEGEGS--LESQPPPLQ----TQACPESSCLREGEKGQNGDDSSAGG--DFP-PPAEVE 110
Db        114 SFEDELSEVFENQSPAAKHGDAASEAPSKDTVEKREDSDKGQQDAFEGTTEGPRPQAFPE 173

Qy        111 PTPEAELL--AQPCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPS--KKKRHWKP 166
Db        174 PKQESSMMGNSQSPGEDTANNTQSPTSLPSQEHGIPQTTEGSE---RGPSAQQQARKAKQ 230

Qy        167 YYKLTWEEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNT------TQFLMDDHDQEEPDL 220
Db        231 EEKEEEEEEEKEEEEEEKEEKAIAREKAGPKEVPTAASSSHFYSGYKKIQKDDDGQSESQA 290

Qy        221 KTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSDGM---------GGDGSEFLQRDFSETY 271
Db        291 VNG---KTGASEAVPSEGKGELEHSQQEEDGEEAMAGPPQGLFPGGKGQE-LERKQQEEE 346

Qy        272 ERYHTESLQNMSKQELIKEYLELEKCLSRMEDENNRLRLESKRLGGDDARVRELELELDR 331
Db        347 E----------EEERLSREW--EDKRWSRMDQLAKELTAE-KRLEGEDDPDRSMKLSF-R 392

Qy        332 LRA 334
Db        393 ARA 395


RESULT 10
AF6_HUMAN
ID   AF6_HUMAN      STANDARD;      PRT;  1816 AA.
AC   P55196; O75087; O75088; O75089; Q9NU92;
DT   01-OCT-1996 (Rel. 34, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   AF-6 protein.
GN   MLLT4 OR AF6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=94061833; PubMed=8242616;
RA   Prasad R., Gu Y., Alder H., Nakamura T., Canaani O., Saito H.,
RA   Huebner K., Gale R.P., Nowell P.C., Kuriyama K., Miyazaki Y.,
RA   Croce C.M., Canaani E.;
RT   "Cloning of the ALL-1 fusion partner, the AF-6 gene, involved in
RT   acute myeloid leukemias with the t(6;11) chromosome translocation.";
RL   Cancer Res. 53:5624-5628(1993).
RN   [2]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
```

```
RC   TISSUE=Fetal brain;
RX   MEDLINE=98344142; PubMed=9679199;
RA   Saito S., Matsushima M., Shirahama S., Minaguchi T., Kanamori Y.,
RA   Minami M., Nakamura Y.;
RT   "Complete genomic structure, DNA polymorphisms, and alternative
RT   splicing of the human AF-6 gene.";
RL   DNA Res. 5:115-120(1998).
RN   [3]
RP   SEQUENCE OF 337-1816 FROM N.A. (ISOFORM 2).
RA   Williams S.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MAY ACT AS AN INTRACELLULAR SIGNALING COMPONENT
CC       CONTROLLED BY RAS SIGNALING PATHWAYS.
CC   -!- SUBUNIT: BINDS DIRECTLY TO ZO-1 AND OCCLUDIN.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=2;
CC         IsoId=P55196-1; Sequence=Displayed;
CC       Name=1;
CC         IsoId=P55196-2; Sequence=VSP_000217, VSP_000218;
CC       Name=3;
CC         IsoId=P55196-3; Sequence=VSP_000219, VSP_000220;
CC   -!- DISEASE: Involved in acute leukemias by a chromosomal
CC       translocation t(6;11)(q27;q23) that involves MLLT4 and MLL/HRX.
CC       The result is a rogue activator protein.
CC   -!- SIMILARITY: Contains 1 dilute domain.
CC   -!- SIMILARITY: Contains 1 FHA domain.
CC   -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC   -!- SIMILARITY: Contains 2 Ras-associating domains.
CC   -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC       WWW="http://www.infobiogen.fr/services/chromcancer/Genes/AF6.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB011399; BAA32484.1; -.
DR   EMBL; AB011399; BAA32483.1; -.
DR   EMBL; AB011399; BAA32485.1; -.
DR   EMBL; U02478; AAC50059.1; -.
DR   EMBL; AL049698; CAB76850.1; -.
DR   HSSP; Q12923; 3PDZ.
DR   Genew; HGNC:7137; MLLT4.
DR   MIM; 159559; -.
DR   GO; GO:0005911; C:intercellular junction; TAS.
DR   GO; GO:0008022; F:protein C-terminus binding activity; TAS.
DR   GO; GO:0007155; P:cell adhesion; TAS.
DR   GO; GO:0007267; P:cell-cell signaling; TAS.
DR   GO; GO:0007048; P:oncogenesis; TAS.
DR   InterPro; IPR002710; DIL.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR000159; RA_domain.
DR   Pfam; PF01843; DIL; 1.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00595; PDZ; 1.
DR   Pfam; PF00788; RA; 2.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00228; PDZ; 1.
DR   SMART; SM00314; RA; 2.
DR   PROSITE; PS50106; PDZ; 1.
DR   PROSITE; PS50200; RA; 2.
KW   Chromosomal translocation; Proto-oncogene; Alternative splicing;
KW   Repeat.
FT   DOMAIN       39    133       RAS-ASSOCIATING 1.
FT   DOMAIN      245    347       RAS-ASSOCIATING 2.
FT   DOMAIN      425    491       FHA.
FT   DOMAIN      804    910       DILUTE.
FT   DOMAIN      991   1077       PDZ.
FT   DOMAIN      162    174       GLU/LYS-RICH.
FT   DOMAIN     1349   1356       POLY-PRO.
FT   DOMAIN     1371   1376       POLY-PRO.
FT   DOMAIN     1561   1571       ASP/GLU-RICH (ACIDIC).
FT   SITE         26     26       MLL FUSION POINT
FT                                (IN AN ACUTE MYELOID LEUKEMIA PATIENT).
FT   VARSPLIC   1588   1611       LQDEERRRQQQLEEMRKREAEDRA -> VKGGVLWLCPSVV
FT                                PILASACFPWG (in isoform 1).
FT                                /FTId=VSP_000217.
FT   VARSPLIC   1612   1816       Missing (in isoform 1).
FT                                /FTId=VSP_000218.
FT   VARSPLIC   1666   1743       LCRPPLPRDYEPPSPSPAPGAPPPPPQRNASYLKTQVLSPD
FT                                SLFTAKFVAYNEEEEEEDCSLAGQDKYSSTRKSHGDL ->
FT                                PNSYPGSTGAAVGAHDACRDAKEKRSKSQDADSPGSSGAPE
FT                                NLTFKERQRLFSQGQDVSNKVKASRKLTELENELNTK (in
FT                                isoform 3).
FT                                /FTId=VSP_000219.
FT   VARSPLIC   1744   1816       Missing (in isoform 3).
FT                                /FTId=VSP_000220.
FT   CONFLICT    373    373       G -> V (IN REF. 1).
FT   CONFLICT    391    391       P -> PGRRNHFAYYNYHTYE (IN REF. 3).
FT   CONFLICT    744    744       D -> DSSHFD (IN REF. 3).
FT   CONFLICT   1031   1031       D -> DV (IN REF. 1).
FT   CONFLICT   1408   1408       R -> P (IN REF. 1).
SQ   SEQUENCE   1816 AA;  205604 MW;  EB1FE7F04879CE8F CRC64;

  Query Match             6.8%;  Score 129.5;  DB 1;  Length 1816;
  Best Local Similarity   20.3%;  Pred. No. 5.5;
  Matches   78;  Conservative   51;  Mismatches  144;  Indels  111;  Gaps   14;

Qy         14 QTSNCTGAAAVQEELN--PERPPGAEERVPEEDSRWQS-RAFPQLGGRPGPEGEGSLESQ 70
              |:|:   : : || ||   :    |          ||::  | |               |
Db       1291 QSSSLDSSTSSQEHLNHSSKSVTPASTLTKSGPGRWKTPAAIPATPVAVSQPIRTDLPPP 1350

Qy         71 PPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELLAQPCHDSEASKL 130
              |||                    ||      | | |    | |             |:::
Db       1351 PPPPPVHYA-------------GDFDGMSMDLPLP----PPP-----------SANQI 1380

Qy        131 GAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWE-EKKKFDEKQSLRASRI 189
              | |:|           |:   : |:|   :: : |   :|  | |:|: :::: |   |
Db       1381 GLPSA----------QVAAAERRKREEHQRWYEKEKARLEEERERKRREQERKLGQMRT 1429

Qy        190 RAEMFAKGQPVAPYNTTQF--------------LMDDHDQEEP------DLKTGLYSKR 228
              ::      | :|    |                : :   |::|      ||:    ||
Db       1430 QS---LNPAPFSPLTAQQMKPEKPSTLQRPQETVIRELQPQQQPRTIERRDLQYITVSKE 1486

Qy        229 AAAKSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELI 288
              :  |  | | :                  :||  |  |:     : :|    |
Db       1487 ELSSGDSLSPDPW---------------------KRDAKEKLEKQQQMHIVDM----LS 1520

Qy        289 KEYLELEKCLSRMEDENNRLR-----------LESKRLGGDDARVRELELE----LDRL 332
              ||  ||:    |  :|::|||            |||:   ||    : :::    : ||
Db       1521 KEIQELQSKPDRSAEESDRLRKLMLEWQFQKRLQESKQKDEDDEEEEDDDVDTMLIMQRL 1580

Qy        333 RAENLQLLTENELHRQQERAPLSK 356
               ||    | : |  |||:   : |
Db       1581 EAERRARLQDEERRRQQQLEEMRK 1604


RESULT 11
VNUA_PRVKA
ID   VNUA_PRVKA     STANDARD;      PRT;  1733 AA.
AC   P33485;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   Probable nuclear antigen.
OS   Pseudorabies virus (strain Kaplan) (PRV).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
OX   NCBI_TaxID=33703;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91021039; PubMed=2171211;
RA   Vlcek C., Kozmik Z., Paces V., Schirm S., Schwyzer M.;
RT   "Pseudorabies virus immediate-early gene overlaps with an oppositely
RT   oriented open reading frame: characterization of their promoter and
RT   enhancer regions.";
RL   Virology 179:365-377(1990).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M34651; AAA47471.1; -.
DR   PIR; B45344; B45344.
FT   DOMAIN      112    117       POLY-THR.
FT   DOMAIN      179   1733       GLY-RICH.
FT   DOMAIN      192    196       POLY-SER.
FT   DOMAIN      271    298       POLY-PRO.
FT   DOMAIN      304    308       POLY-ARG.
FT   DOMAIN      883    889       POLY-GLY.
FT   DOMAIN     1398   1405       POLY-GLY.
SQ   SEQUENCE   1733 AA;  172166 MW;  0C8CD8BE475BB5E2 CRC64;

  Query Match             6.7%;  Score 128;  DB 1;  Length 1733;
  Best Local Similarity   22.4%;  Pred. No. 6.3;
  Matches   74;  Conservative   27;  Mismatches   94;  Indels  136;  Gaps   16;

Qy          8 EYQHQPQTSNCT---------------GAAAVQEE----LNPER-PPGAEER-------- 39
              ::||:|  :  |                    :|||    | | | || |:
Db        104 DHQHRPPPTTTTTTIKDPQHPQDPLLLPTKTLQEEDPHLLRPTRDPPSAKTHHHQDPPGG 163

Qy         40 VPEEDSRWQSRAFPQLGGRPGP----------EGEGSLESQPPPLQ-------------- 75
               |   |      |  || | |           :|   ::||| |
Db        164 GPPSTSSHHHHQDPPGGGPPSPPPRPSTSSSSSHQGPPSTRPPPPQRPPPRWPPPSPQKI 223

Qy         76 --TQACPESSCLR-------------EGEKGQNGDDSSAGG-----DFPPPAEVEPTPEA 115
                |:|  |::               || |: ||  :|||     | |||    | |
Db        224 SETRAGSENTAQTLFSHSENKLFSHPMGEGGE-GDRGTAGGGEGDRDDPPPPSPPPRPPP 282

Qy        116 ELLAQPCHDSEASKLGAPAAGGEEEWGQQQRQLG---KKKHRRRPSKKKRHWKPYYKLTW 172
               |   |           | |||     :::|: |    :  |||  |::|
Db        283 PLPPPPPPPPPPQ---PPPAGGS---ARRRRRGGGPPGRGGRRRGGKRRR---------- 326

Qy        173 EEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAK 232
                                |:|   |          |  ::|: |
Db        327 ---------------------AEGTEAA-------AADAEEEEDGD------------- 344

Qy        233 SDDTSDDDFMEEGGEEDGGSD--GMGGDGSE 261
              |:  |:|  |  | ||||    | ||   |
Db        345 GDEDEDEDRAEGEGREDGGEGPRGAGGGAGE 375


RESULT 12
ATRX_MOUSE
ID   ATRX_MOUSE     STANDARD;      PRT;  2476 AA.
AC   Q61687;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Transcriptional regulator ATRX (X-linked nuclear protein)
DE   (Heterochromatin protein 2) (HP1 alpha-interacting protein) (HP1-BP38
DE   protein).
GN   ATRX OR XNP OR HP1BP2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98213653; PubMed=9545503;
RA   Picketts D.J., Tastan A.O., Higgs D.R., Gibbons R.J.;
RT   "Comparison of the human and murine ATRX gene identifies highly
RT   conserved, functionally important domains.";
RL   Mamm. Genome 9:400-403(1998).
RN   [2]
RP   SEQUENCE OF 325-1176 FROM N.A.
RX   MEDLINE=97133299; PubMed=8978696;
RA   le Douarin B., Nielsen A.L., Garnier J.-M., Ichinose H.,
RA   Jeanmougin F., Losson R., Chambon P.;
RT   "A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic
RT   control of transcription by nuclear receptors.";
RL   EMBO J. 15:6701-6715(1996).
RN   [3]
RP   SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
RP   HETEROCHROMATIN.
RX   MEDLINE=20040663; PubMed=10570185;
RA   McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
RA   Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,
RA   Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
RT   "Localization of a putative transcriptional regulator (ATRX) at
RT   pericentromeric heterochromatin and the short arms of acrocentric
RT   chromosomes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
CC   -!- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
CC       GENE EXPRESSION BY AFFECTING CHROMATIN.
CC   -!- SUBUNIT: PROBABLY BINDS EZH2. BINDS ANNEXIN V IN A CALCIUM AND
CC       PHOSPHATIDYLCHOLINE/PHOSPHATIDYLSERINE-DEPENDENT MANNER (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC
CC       HETEROCHROMATIN DURING INTERPHASE AND MITOSIS, PROBABLY BY
CC       INTERACTING WITH HP1.
CC   -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC   -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF026032; AAC08741.1; -.
DR   EMBL; X99643; CAA67962.1; -.
DR   MGD; MGI:103067; Atrx.
DR   GO; GO:0000228; C:nuclear chromosome; IDA.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding;
KW   Zinc-finger.
FT   ZN_FING     219    267       PHD-TYPE.
FT   NP_BIND    1579   1586       ATP (POTENTIAL).
FT   SITE       1704   1707       DEGH BOX.
FT   DOMAIN      319    322       POLY-SER.
FT   DOMAIN      735    738       POLY-SER.
FT   DOMAIN     1001   1004       POLY-GLU.
FT   DOMAIN     1130   1135       POLY-SER.
FT   DOMAIN     1182   1185       POLY-SER.
FT   DOMAIN     1238   1245       POLY-ASP.
FT   DOMAIN     1484   1487       POLY-GLU.
FT   DOMAIN     1924   1931       POLY-SER.
FT   DOMAIN     2205   2208       POLY-LYS.
FT   DOMAIN     2245   2248       POLY-GLU.
FT   DOMAIN     2403   2408       POLY-GLN.
SQ   SEQUENCE   2476 AA;  278601 MW;  90A42B790FC4FF4C CRC64;

  Query Match             6.7%;  Score 128;  DB 1;  Length 2476;
```

```
  Best Local Similarity   19.7%;  Pred. No. 9.1;
  Matches   78;  Conservative   52;  Mismatches  130;  Indels  136;  Gaps   13;

Qy          43 EDSRWQSRAFPQLGGRPGPEGE---GSLESQPPPLQTQACPESSCLREGEKGQNGDDSSA 99
               ||:: | :       : |   |    || : |   |       || : : ::   |::||
Db        1142 EDNKKQKKQRTSAKKKTGNTKEKKRNSLRATPKRKQVDITSSSSDIGDDDQNSAGEESSD 1201

Qy         100 GGDFPP-------------------------PAEVEP-----TPE---------AELLA 119
                 |                            || |:       ||          |: |
Db        1202 EQKIKPVTENLVLPSHTGFCQSSGDEALSKSVPATVDDDDDDNDPENRIAKKMLLEEIKA 1261

Qy         120 QPCHDSEASKLGAPAAGGEEEWGQQQ-----------------RQLGKKKHRRRPSKKK- 161
                 | : |    |  ||::  |:|                  |:  :          ||
Db        1262 NLSSDEDGSSDDEPDGGGKKRIGKQSEESPADDGELRREQLAVNQVNSESDSDSEESKKP 1321

Qy         162 --RHWKPYYKLTWEEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPD 219
               ||    :|||  : :  :||     :: :    |||:
Db        1322 RYRHRLLRHKLTLSDGESGEEK----PTKPKEHKEAKGR---------------------- 1356

Qy         220 LKTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESL 279
                   |    |:|: | || | |  |:           ||  ||  : : ::   |
Db        1357 -------NRRKVSSEDSEDTDFQESGVSEEVSE-------SEDEQRPRTRSAKKAELEEN 1402

Qy         280 QNMSKQELIKEYLELEKCLSR-----------------------MEDENNRLRLESKRL 315
               |   ||:  :  :::::  |                      ||||:    :||
Db        1403 QRSYKQKKKRRRIKVQEDSSSENKSHSEEDKKEGDEEDEEDEDEDEEDEND----DSKSP 1458

Qy         316 GGDDARVRELELELDRLRAENLQLLTENELHRQQER 351
               |    ::|:: |: |:|| |     |:| |  ::||
Db        1459 GKGRKKIRKI-LKDDKLRTE-----TQNALKEEEER 1488
```

```
RESULT 13
PLE1_RAT
ID   PLE1_RAT       STANDARD;      PRT;  4687 AA.
AC   P30427; O08879; O08880; O08881;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Plectin 1 (PLTN) (PCN).
GN   PLEC1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Glial tumor;
RX   MEDLINE=91268156; PubMed=2050743;
RA   Wiche G., Becker B., Luber K., Weitzer G., Castanon M.J.,
RA   Hauptmann R., Stratowa C., Stewart M.;
RT   "Cloning and sequencing of rat plectin indicates a 466-kD polypeptide
RT   chain with a three-domain structure based on a central alpha-helical
RT   coiled coil.";
RL   J. Cell Biol. 114:83-99(1991).
RN   [2]
RP   REVISIONS.
RC   TISSUE=Glial tumor;
RX   MEDLINE=96210632; PubMed=8633055;
RA   Liu C.-G., Maercker C., Castanon M.J., Hauptmann R., Wiche G.;
RT   "Human plectin: organization of the gene, sequence analysis, and
RT   chromosome localization (8q24).";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283(1996).
RN   [3]
RP   PARTIAL SEQUENCE FROM N.A. (ISOFORMS 2; 3 AND 4), AND TISSUE
RP   SPECIFICITY.
RC   TISSUE=Glial tumor;
RX   MEDLINE=97321050; PubMed=9177781;
RA   Elliott C.E., Becker B., Oehler S., Castanon M.J., Hauptmann R.,
RA   Wiche G.;
RT   "Plectin transcript diversity: identification and tissue distribution
RT   of variants with distinct first coding exons and rodless isoforms.";
RL   Genomics 42:115-125(1997).
```

```
CC   -!- FUNCTION: INTERLINKS INTERMEDIATE FILAMENTS WITH MICROTUBULES AND
CC       MICROFILAMENTS AND ANCHORS INTERMEDIATE FILAMENTS TO DESMOSOMES OR
CC       HEMIDESMOSOMES. MAY BE INVOLVED NOT ONLY IN THE CROSSLINKING AND
CC       STABILIZATION OF CYTOSKELETAL INTERMEDIATE FILAMENTS NETWORK, BUT
CC       ALSO IN THE REGULATION OF THEIR DYNAMICS.
CC   -!- SUBUNIT: HOMODIMER OR HOMOTETRAMER.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=4;
CC       Name=1;
CC         IsoId=P30427-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=P30427-2; Sequence=VSP_005050;
CC       Name=3;
CC         IsoId=P30427-3; Sequence=VSP_005051;
CC       Name=4;
CC         IsoId=P30427-4; Sequence=VSP_005052;
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST EXPRESSION IN
CC       SKELETAL MUSCLE AND LOWEST IN THYMUS.
CC   -!- DOMAIN: THE N-TERMINUS INTERACTS WITH ACTIN, THE C-TERMINUS WITH
CC       VIMENTIN, DESMIN, GFAP, CYTOKERATINS, LAMIN B; WHEREAS BOTH THE N-
CC       AND THE C-TERMINUS CAN BIND INTEGRIN BETA-4.
CC   -!- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM
CC       INTERMEDIATE FILAMENTS DURING MITOSIS (BY SIMILARITY).
CC   -!- SIMILARITY: Contains 1 actin-binding domain.
CC   -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC   -!- SIMILARITY: Contains 33 plectin repeats.
CC   -!- SIMILARITY: Contains 4 spectrin repeats.
CC   -!- SIMILARITY: Belongs to the plakin or cytolinker family.
CC   -!- CAUTION: ISOFORM 4 IS A FRAGMENT AT THE N-TERMINUS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X59601; CAA42169.1; -.
DR   EMBL; U96274; AAC53209.1; -.
DR   EMBL; U96275; AAC53210.1; -.
DR   EMBL; U96276; AAC53211.1; -.
DR   PIR; A39638; A39638.
DR   HSSP; Q01082; 1BKR.
DR   InterPro; IPR001589; Actbind_actnin.
DR   InterPro; IPR001715; Calponin-like.
DR   InterPro; IPR001101; Plectin_repeat.
DR   InterPro; IPR005326; S10_plectin.
DR   InterPro; IPR002017; Spectrin.
DR   Pfam; PF00307; CH; 2.
DR   Pfam; PF00681; Plectin; 21.
DR   Pfam; PF03501; S10_plectin; 1.
DR   ProDom; PD006662; S10_plectin_N; 1.
DR   SMART; SM00033; CH; 2.
DR   SMART; SM00250; PLEC; 33.
DR   SMART; SM00150; SPEC; 6.
DR   PROSITE; PS00019; ACTININ_1; FALSE_NEG.
DR   PROSITE; PS00020; ACTININ_2; FALSE_NEG.
DR   PROSITE; PS50021; CH; 2.
KW   Coiled coil; Repeat; Structural protein; Cytoskeleton; Actin-binding;
KW   Phosphorylation; Alternative splicing.
FT   DOMAIN        1   1473       GLOBULAR 1.
FT   DOMAIN     1474   2758       CENTRAL FIBROUS ROD DOMAIN.
FT   DOMAIN     2759   4687       GLOBULAR 2.
FT   DOMAIN      181    406       ACTIN-BINDING.
FT   DOMAIN      185    288       CH 1.
FT   DOMAIN      301    403       CH 2.
FT   REPEAT      648    722       SPECTRIN 1.
FT   REPEAT      743    827       SPECTRIN 2.
FT   REPEAT      840    933       SPECTRIN 3.
FT   REPEAT     1318   1418       SPECTRIN 4.
FT   DOMAIN     1472   1692       COILED COIL (POTENTIAL).
FT   DOMAIN     1724   2760       COILED COIL (POTENTIAL).
```

```
FT   REPEAT     2791   2828       PLECTIN 1.
FT   REPEAT     2829   2866       PLECTIN 2.
FT   REPEAT     2867   2904       PLECTIN 3.
FT   REPEAT     2905   2942       PLECTIN 4.
FT   REPEAT     2943   2980       PLECTIN 5.
FT   REPEAT     2984   3018       PLECTIN 6.
FT   REPEAT     3119   3156       PLECTIN 7.
FT   REPEAT     3157   3194       PLECTIN 8.
FT   REPEAT     3195   3232       PLECTIN 9.
FT   REPEAT     3233   3270       PLECTIN 10.
FT   REPEAT     3271   3308       PLECTIN 11.
FT   REPEAT     3311   3346       PLECTIN 12.
FT   REPEAT     3488   3525       PLECTIN 13.
FT   REPEAT     3526   3563       PLECTIN 14.
FT   REPEAT     3564   3601       PLECTIN 15.
FT   REPEAT     3602   3639       PLECTIN 16.
FT   REPEAT     3643   3677       PLECTIN 17.
FT   REPEAT     3823   3860       PLECTIN 18.
FT   REPEAT     3861   3898       PLECTIN 19.
FT   REPEAT     3899   3936       PLECTIN 20.
FT   REPEAT     3937   3974       PLECTIN 21.
FT   REPEAT     3978   4011       PLECTIN 22.
FT   REPEAT     4066   4103       PLECTIN 23.
FT   REPEAT     4104   4141       PLECTIN 24.
FT   REPEAT     4142   4179       PLECTIN 25.
FT   REPEAT     4180   4217       PLECTIN 26.
FT   REPEAT     4221   4255       PLECTIN 27.
FT   REPEAT     4268   4308       PLECTIN 28.
FT   REPEAT     4411   4448       PLECTIN 29.
FT   REPEAT     4449   4486       PLECTIN 30.
FT   REPEAT     4487   4524       PLECTIN 31.
FT   REPEAT     4525   4562       PLECTIN 32.
FT   REPEAT     4563   4600       PLECTIN 33.
FT   DOMAIN     4253   4303       BINDING TO INTERMEDIATE FILAMENTS.
FT   DOMAIN     4628   4643       4 X 4 AA TANDEM REPEATS OF G-S-R-X.
FT   MOD_RES    4542   4542       PHOSPHORYLATION (BY CDC2)
FT                                (BY SIMILARITY).
FT   VARSPLIC      1    180       MVAGMLMPLDQLRAIYEVLFREGVMVAKKDRRPRSLHPHVP
FT                                GVTNLQVMRAMTSLKARGLVRETFAWCHFYWYLTNEGIDHL
FT                                RQYLHLPPEIVPASLQRVRRPVAMVMPARRRSPHVQTMQGP
FT                                LGCPPKRGPLPAEDPAREERQVYRRKEREEGAPETPVVSAT
FT                                IVGTLARPGPEPTPAT -> MSQQRLRVPEPEGLGSKRTSS
FT                                EDNLYLAVLRASEGKK (in isoform 2).
FT                                /FTId=VSP_005050.
FT   VARSPLIC      1    180       MVAGMLMPLDQLRAIYEVLFREGVMVAKKDRRPRSLHPHVP
FT                                GVTNLQVMRAMTSLKARGLVRETFAWCHFYWYLTNEGIDHL
FT                                RQYLHLPPEIVPASLQRVRRPVAMVMPARRRSPHVQTMQGP
FT                                LGCPPKRGPLPAEDPAREERQVYRRKEREEGAPETPVVSAT
FT                                IVGTLARPGPEPTPAT -> MEPSGSLFPSLVVVGHVVSLA
FT                                AVWHWRKGHRQAQDEQ (in isoform 3).
FT                                /FTId=VSP_005051.
FT   VARSPLIC      1    180       MVAGMLMPLDQLRAIYEVLFREGVMVAKKDRRPRSLHPHVP
FT                                GVTNLQVMRAMTSLKARGLVRETFAWCHFYWYLTNEGIDHL
FT                                RQYLHLPPEIVPASLQRVRRPVAMVMPARRRSPHVQTMQGP
FT                                LGCPPKRGPLPAEDPAREERQVYRRKEREEGAPETPVVSAT
FT                                IVGTLARPGPEPTPAT -> DVSNGSSGSPSPGDTLPWNLG
FT                                KTQRSRRSGGGSVGNGSVLDPAERAVIRIA (in
FT                                isoform 4).
FT                                /FTId=VSP_005052.
SQ   SEQUENCE   4687 AA;  533527 MW;  9966CAF71B929751 CRC64;

  Query Match             6.7%;  Score 127.5;  DB 1;  Length 4687;
  Best Local Similarity   21.4%;  Pred. No. 19;
  Matches   78;  Conservative   51;  Mismatches  127;  Indels  109;  Gaps   13;

Qy         25 QEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPGPEGEGSLESQPPPLQTQACPESSC 84
              |::   ||   | |    |  ||| :|   |  |              || :   :
Db       1755 QQQAEAER---AREEAERELERWQLKANEALRLR--------------LQAEEVAQQKS 1796

Qy         85 LREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELLAQPCH---------------DSEAS 128
              | :  :  :  :::          || :   : ||  |                 : |
Db       1797 LAQADAEKQKEEAEREARRRGKAEEQAVRQRELAEQELEKQRQLTEGTAQQRLAAEQELI 1856

Qy        129 KLGAPAAGGEEEWGQQQRQ-----LGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDEKQS 183
              :| |     | | |: |||     | : :|    : :||      :|  |  |   | :
Db       1857 RLRA-----ETEQGEHQRQLLEEELARLQHEATAATQKRQ-----ELEAELAKVRAEMEV 1906

Qy        184 LRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRA-----------AA 231
              | ||: |||  ::               :  ::  : : | : : |           |
Db       1907 LLASKARAEEESRS-----------TSEKSKQRLEAEAGRFRELAEEAARLRALAEEAR 1954

Qy        232 KSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEY 291
              :  : :::|   :  | ||                              || |  :|:  :|
Db       1955 RHRELAEEDAARQRAEADG----------------------VLTEKLAAISEATRLK-- 1989

Qy        292 LELEKCLSRMEDENNRLRLESKRLGGDDA-RVRELELELDRLRAE------NLQLLTENE 344
              | |  |   | || |||    ||  |:| : | || :  : :|:       |:  :|:|
Db       1990 TEAEIALKEKEAENERLR----RLAEDEAFQRRRLEEQAAQHKADIEERLAQLRKASESE 2045

Qy        345 LHRQQ 349
              | ||:
Db       2046 LERQK 2050


RESULT 14
TRHY_RABIT
ID   TRHY_RABIT     STANDARD;      PRT;  1407 AA.
AC   P37709;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Trichohyalin.
GN   THH.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fietz M.J., Rogers G.E.;
RL   Submitted (DEC-1992) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
CC       IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
CC       INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
CC       LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
CC       ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
CC       WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
CC       ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
CC       ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
CC       DIFFERENTIATION.
CC   -!- SUBUNIT: Homodimer (Probable).
CC   -!- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
CC       THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
CC       THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
CC       THE EPIDERMIS.
CC   -!- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CC       CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
CC       ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
CC       OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED
CC       ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
CC       THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
CC       DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
CC       THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
CC       DIFFERENT SPECIES.
CC   -!- PTM: SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE
CC       PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100
CC       FAMILY.
CC   -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z19092; CAA79519.1; -.
DR   PIR; S28589; S28589.
DR   HSSP; P02633; 4ICB.
DR   InterPro; IPR001751; CaBP_S100.
DR   InterPro; IPR002048; EF-hand.
DR   Pfam; PF00036; efhand; 1.
DR   Pfam; PF01023; S_100; 1.
DR   ProDom; PD003407; CaBP_S100; 1.
DR   PROSITE; PS00018; EF_HAND; 1.
DR   PROSITE; PS00303; S100_CABP; 1.
KW   Keratinization; Calcium-binding; Repeat; Citrullination.
FT   DOMAIN        1     91       S-100 LIKE.
FT   CA_BIND      22     33       EF-HAND 1 (LOW AFFINITY) (POTENTIAL).
FT   CA_BIND      62     73       EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).
SQ   SEQUENCE   1407 AA;  183781 MW;  AE17D2A159F12B7F CRC64;

  Query Match             6.6%;  Score 126.5;  DB 1;  Length 1407;
  Best Local Similarity   23.7%;  Pred. No. 6.1;
  Matches   83;  Conservative   48;  Mismatches  124;  Indels   95;  Gaps   17;

Qy         25 QEE-----LNPERPPGAEERVPEEDSRWQSRAFPQLGGRPGPEGEGSLESQPPPLQTQAC 79
              |||     |  |    | ||     ||| :   : | |     :  : |:|   : |:
Db        349 QEERREQLLAEEVREQARERGESLTRRWQRQLESEAGAR-----QSKVYSRPRRQEEQS- 402

Qy         80 PESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELLAQPCHDSEASKLGAPAAGGEE 139
                 ||: :: :   :                  | ||  |               |  ::
Db        403 -----LRQDQERRQRQER---------------ERELEEQ--------------ARRQQ 427

Qy        140 EW-GQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDEKQSLRASRIRAEMFAKGQ 198
              :|  ::: :  :::   ||| ::|  :   :   |::::| |::  |  | | |:
Db        428 QWQAEEESERRRQRLSARPSLRERQLRAEER--QEQEQRFREEEEQRRER-RQEL----- 479

Qy        199 PVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSDGMGGD 258
                  |||     :||  |      :|       :| | |:       :   |
Db        480 --------QFL-----EEEEQL------QRRERAQQLQEEDSFQEDRERRRRQQEQRPGQ 520

Qy        259 GSEFLQRDFSETYERYHT-------------ESLQ-NMSKQELIKEYLELEKCLSRMED 303
                :  ::  |   | ||               | ||    :||  :|| | || | | ||
Db        521 TWRWQLQE--EAQRRRHTLYAKPGQQEQLREEEELQREKRRQERERREYREEEK-LQREED 577

Qy        304 ENNRLRLESKRLGGDDARVRELELELDRLRAENLQLLTENELH--RQQER 351
              |  | | | :|   :   :|: |   ||   |  ||| | |    |:|||
Db        578 EKRR-RQERERQYRELEELRQEEQLRDRKLREEEQLLQEREEERLRRQER 626


RESULT 15
DGR8_HUMAN
ID   DGR8_HUMAN     STANDARD;      PRT;   773 AA.
AC   Q8WYQ5; Q96G39; Q96GP8; Q9H6L8; Q9H6T7; Q9NRW2;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   DGCR8 protein (DiGeorge syndrome critical region 8).
GN   DGCR8 OR DGCRK6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RA   Shimizu N., Minosima S., Kawasaki K., Sasaki T., Shiohama A.;
RT   "Homo sapiens DGCRK6 on 22q11.2.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   TISSUE=Brain, and Muscle;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 1-247 FROM N.A.
RC   TISSUE=Heart;
RA   Gong L., Millas S., Jen J., Yeh E.T.H.;
RT   "Isolation and characterization of a novel human gene deleted in
RT   DiGeorge syndrome.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 204-773 FROM N.A. (ISOFORM 1).
RC   TISSUE=Hepatoma;
RA   Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA   Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA   Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q8WYQ5-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q8WYQ5-2; Sequence=VSP_003847, VSP_003848;
CC         Note=No experimental confirmation available;
CC   -!- DISEASE: May play a part in the etiology of the
CC       velocardiofacial/DiGeorge syndrome (VCFS/DGS), a developmental
CC       disorder characterized by structural and functional palate
CC       anomalies, conotruncal cardiac malformations, immunodeficiency,
CC       hypocalcemia, and typical facial anomalies. Most cases result from
CC       a deletion of chromosome 22q11.2 (the DiGeorge syndrome chromosome
CC       region, or DGCR).
CC   -!- SIMILARITY: Contains 2 DRBM (double-stranded RNA-binding) domains.
CC   -!- SIMILARITY: Contains 1 WW domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AB050770; BAB83032.1; -.
DR   EMBL; BC009984; AAH09984.1; -.
DR   EMBL; BC009323; AAH09323.1; ALT_INIT.
DR   EMBL; AF165527; AAF82263.1; -.
DR   EMBL; AK025539; BAB15165.1; ALT_INIT.
DR   EMBL; AK025780; BAB15238.1; ALT_INIT.
DR   Genew; HGNC:2847; DGCR8.
DR   InterPro; IPR001159; DS_RBD.
DR   InterPro; IPR001202; WW_Rsp5_WWP.
DR   Pfam; PF00035; dsrm; 2.
DR   Pfam; PF00397; WW; 1.
DR   SMART; SM00358; DSRM; 2.
DR   SMART; SM00456; WW; 1.
DR   PROSITE; PS50137; DS_RBD; 1.
DR   PROSITE; PS01159; WW_DOMAIN_1; FALSE_NEG.
DR   PROSITE; PS50020; WW_DOMAIN_2; 1.
```

```
KW   RNA-binding; Repeat; Alternative splicing.
FT   DOMAIN      301    334       WW.
FT   DOMAIN      511    578       DRBM 1.
FT   DOMAIN      620    685       DRBM 2.
FT   VARSPLIC    303    304       LP -> VL (in isoform 2).
FT                                /FTId=VSP_003847.
FT   VARSPLIC    305    773       Missing (in isoform 2).
FT                                /FTId=VSP_003848.
FT   CONFLICT    241    247       DDFDNDV -> VCWQPLL (IN REF. 3).
FT   CONFLICT    255    256       LC -> TR (IN REF. 2; AAH09323).
FT   CONFLICT    274    274       P -> L (IN REF. 4).
FT   CONFLICT    343    343       H -> Y (IN REF. 4).
FT   CONFLICT    706    706       V -> A (IN REF. 4; BAB15165).
SQ   SEQUENCE   773 AA;  86045 MW;  72D962BBE32890EC CRC64;

  Query Match             6.6%;  Score 125.5;  DB 1;  Length 773;
  Best Local Similarity   23.1%;  Pred. No. 3.7;
  Matches   74;  Conservative   41;  Mismatches  108;  Indels   97;  Gaps   15;

Qy         58 RPGPEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPP------PAEVE- 110
              ||   | ||:    ||| :  |     ::: |: :   | :  ||  |       ||:|
Db        331 RPYFLGTGSIRKHDPPLSSIPCLHYKKMKDNEEREQSSDLTPSGDVSPVKPLSRSAELEF 390

Qy        111 PTPEAELL-AQPCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKK----RHWK 165
              |  | : : | |    |   ||| || |          ||: | :    | :      ::
Db        391 PLDEPDSMGADPGPPDEKDPLGAEAAPG---------ALGQVKAKVEVCKDESVDLEEFR 441

Qy        166 PYYK------------LTWEEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDD 212
              | :              || |:::|: :      | :||     :|: | |  | |:
Db        442 SYLEKRFDFEQVTVKKFRTWAERRQFNREM----KRKQAE---SERPILPAN--QKLITL 492

Qy        213 HDQEEPDLKTGLYSKRAAAKSDDTSDDDFME-------------EGGEE--------DG 250
              |: |  |  : :     ||:     ::|:               |   |         ||
Db        493 SVQDAPTKKEFVINPN--GKSEVCILHEYMQRVLKVRPVYNFFECENPSEPFGASVTIDG 550

Qy        251 GSDGMGGDGS------------EFLQRDF----------SETYERYHTESLQN----- 281
              : | |   |              | |  ||          ||  | ::  |:::
Db        551 VTYGSGTASSKKLAKNKAARATLEILIPDFVKQTSEEKPKDSEELEYFNHISIEDSRVYE 610

Qy        282 -MSKQELIKEYLELEKCLSR 300
               ||  |:  |  | :|| |
Db        611 LTSKAGLLSPYQILHECLKR 630


Search completed: February  5, 2004, 13:29:01
Job time : 26 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 5, 2004, 15:09:51 ; Search time 488 Seconds
(without alignments)
2709.886 Million cell updates/sec

Title: US-09-972-758A-2
Perfect score: 1910
Sequence: 1 MAEPFLSEYQHQPQTSNCTG.....LTENELHRQQERAPLSKFGD 359

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09972758/runat_05022004_095005_1309/app_query.fasta_1.519
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09972758_@CGN_1_1_391_@runat_05022004_095005_1309
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1910	100.0	2199	9	US-09-745-763-33	Sequence 33, Appl
2	1910	100.0	3624	15	US-10-205-823-171	Sequence 171, App
3	1902	99.6	1080	10	US-09-972-758-1	Sequence 1, Appli
c 4	733	38.4	1083	12	US-10-264-237-629	Sequence 629, App
5	691	36.2	414	10	US-09-983-965-4895	Sequence 4895, Ap
6	629	32.9	461	10	US-09-954-631-813	Sequence 813, App
c 7	619	32.4	495	11	US-09-918-995-403	Sequence 403, App
8	556.5	29.1	1330	13	US-10-094-749-1223	Sequence 1223, Ap
c 9	545	28.5	2048	12	US-10-108-260A-1496	Sequence 1496, Ap
10	184	9.6	549	11	US-09-991-936-1712	Sequence 1712, Ap
11	162	8.5	6354	15	US-10-084-817-158	Sequence 158, App
12	162	8.5	7396	12	US-10-028-248A-35	Sequence 35, Appl
13	162	8.5	7396	12	US-10-107-782-35	Sequence 35, Appl
14	159.5	8.4	4041	12	US-10-264-049-543	Sequence 543, App
15	155	8.1	2901	12	US-10-085-117-321	Sequence 321, App
16	155	8.1	5002	12	US-10-085-117-320	Sequence 320, App
17	152	8.0	2421	10	US-09-974-300-7891	Sequence 7891, Ap
18	147.5	7.7	3773	9	US-09-925-302-47	Sequence 47, Appl
19	147.5	7.7	4913	12	US-10-159-563-34	Sequence 34, Appl
20	147.5	7.7	4913	13	US-10-133-937-34	Sequence 34, Appl
21	147	7.7	7453	13	US-10-117-722-248	Sequence 248, App
22	147	7.7	7453	15	US-10-037-270-248	Sequence 248, App
23	147	7.7	7501	13	US-10-117-722-249	Sequence 249, App
24	147	7.7	7501	15	US-10-037-270-249	Sequence 249, App
25	146.5	7.7	2479	12	US-10-108-260A-1966	Sequence 1966, Ap
26	145	7.6	4301	10	US-09-917-800A-1666	Sequence 1666, Ap
27	145	7.6	4301	13	US-10-205-194-126	Sequence 126, App
28	144	7.5	2090	13	US-10-094-749-1350	Sequence 1350, Ap
29	144	7.5	4714	12	US-09-825-489-12	Sequence 12, Appl
30	142.5	7.5	1799	12	US-10-108-260A-1885	Sequence 1885, Ap
31	141	7.4	2907	12	US-10-085-117-324	Sequence 324, App
32	141	7.4	2962	11	US-09-291-417-106	Sequence 106, App
33	141	7.4	4221	12	US-10-085-117-323	Sequence 323, App
34	141	7.4	6202	13	US-10-120-988-120	Sequence 120, App
35	140	7.3	5973	13	US-10-094-466-63	Sequence 63, Appl
36	139	7.3	1611	12	US-10-136-728-37	Sequence 37, Appl
37	139	7.3	1691	12	US-10-136-728-35	Sequence 35, Appl
38	139	7.3	6537	15	US-10-175-523-186	Sequence 186, App
39	138.5	7.3	5835	11	US-09-927-597-1	Sequence 1, Appli
40	138.5	7.3	5937	11	US-09-927-597-3	Sequence 3, Appli
41	138.5	7.3	6861	13	US-10-341-434-102	Sequence 102, App
42	138.5	7.3	6861	15	US-10-171-311-161	Sequence 161, App
43	138.5	7.3	6900	15	US-10-171-311-163	Sequence 163, App
44	138	7.2	1960	10	US-09-834-975-833	Sequence 833, App
45	138	7.2	2093	12	US-10-104-047-1666	Sequence 1666, Ap

ALIGNMENTS

RESULT 1
US-09-745-763-33
; Sequence 33, Application US/09745763
; Patent No. US20020065394A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; McCoy, John M.
; LaVallie, Edward R.
; Collins-Racie, Lisa A.
; Evans, Cheryl
; Merberg, David
; Treacy, Maurice
; Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; ENCODING THEM
; NUMBER OF SEQUENCES: 219
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:

```
;              MEDIUM TYPE: Floppy disk
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/09/745,763
;              FILING DATE: 18-Jun-2000
;              CLASSIFICATION: <Unknown>
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Sprunger, Suzanne A.
;              REGISTRATION NUMBER: 41,323
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (617) 498-8284
;              TELEFAX: (617) 876-5851
;    INFORMATION FOR SEQ ID NO: 33:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 2199 base pairs
;              TYPE: nucleic acid
;              STRANDEDNESS: double
;              TOPOLOGY: linear
;         MOLECULE TYPE: cDNA
;         SEQUENCE DESCRIPTION: SEQ ID NO: <Unknown>
US-09-745-763-33

Alignment Scores:
Pred. No.:                1.49e-164       Length:          2199
Score:                    1910.00         Matches:         359
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       9               Gaps:            0

US-09-972-758A-2 (1-359) x US-09-745-763-33 (1-2199)

Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        707 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 766

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        767 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 826

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        827 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 886

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        887 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 946

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        947 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 1006

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1007 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 1066

Qy        121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1067 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 1126

Qy        141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1127 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 1186

Qy        161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1187 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 1246

Qy        181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1247 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 1306

Qy        201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1307 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 1366

Qy        221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1367 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 1426

Qy        241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1427 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 1486

Qy        261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 1546

Qy        281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArg 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 1606

Qy        301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 1666

Qy        321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1726

Qy        341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGAC 1783

RESULT 2
US-10-205-823-171
; Sequence 171, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
;  APPLICANT: Schlegel, Robert
;  APPLICANT:  Monahan, John E.
;  APPLICANT:  Endege, Wilson O.
;  APPLICANT:  Gannavarapu, Manjula
;  APPLICANT:  Gorbatcheva, Bella
;  APPLICANT:  Hoersch, Sebastian
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Wonsey, Angela M.
;  APPLICANT:  Glatt, Karen
;  APPLICANT:  Zhao, Xumei
;  APPLICANT:  Anderson, Dustin
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
;  TITLE OF INVENTION:  METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF PROSTATE CANCER
;  FILE REFERENCE: MRI-044
;  CURRENT APPLICATION NUMBER: US/10/205,823
;  CURRENT FILING DATE:  2002-07-25
;  PRIOR APPLICATION NUMBER: 60/307,982
;  PRIOR FILING DATE: 2001-07-25
;  PRIOR APPLICATION NUMBER: 60/314,356
;  PRIOR FILING DATE: 2001-08-22
;  PRIOR APPLICATION NUMBER: 60/325,020
;  PRIOR FILING DATE: 2001-09-25
;  PRIOR APPLICATION NUMBER: 60/341,746
;  PRIOR FILING DATE: 2001-12-12
;  PRIOR APPLICATION NUMBER: 60/362,158
;  PRIOR FILING DATE: 2002-03-05
;  NUMBER OF SEQ ID NOS: 455
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 171
;   LENGTH: 3624
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-205-823-171
```

Alignment Scores:

Pred. No.:	2.62e-164	Length:	3624
Score:	1910.00	Matches:	359
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	15	Gaps:	0

US-09-972-758A-2 (1-359) x US-10-205-823-171 (1-3624)

```
Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        690 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 749

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        750 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 809

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        810 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 869

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        870 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 929

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        930 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 989

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        990 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 1049

Qy        121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1050 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 1109

Qy        141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1110 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 1169

Qy        161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1170 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 1229

Qy        181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1230 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 1289

Qy        201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1290 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 1349

Qy        221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1350 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 1409

Qy        241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1410 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 1469

Qy        261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1470 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 1529

Qy        281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArg 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1530 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 1589

Qy        301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1590 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 1649

Qy        321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1650 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1709

Qy        341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1710 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGAC 1766
```

RESULT 3
US-09-972-758-1
; Sequence 1, Application US/09972758
; Patent No. US20020160497A1
; GENERAL INFORMATION:
; APPLICANT: Case Western Reserve University
; APPLICANT: Montano, Monica
; APPLICANT: Wittman, Bryan
; TITLE OF INVENTION: Suppressors of Human Breast Cancer Cell Growth
; FILE REFERENCE: 27708/04004
; CURRENT APPLICATION NUMBER: US/09/972,758
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: US 60/238,187
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1080
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-972-758-1

Alignment Scores:

Pred. No.:	3.59e-164	Length:	1080
Score:	1902.00	Matches:	357
Percent Similarity:	99.72%	Conservative:	1
Best Local Similarity:	99.44%	Mismatches:	1
Query Match:	99.58%	Indels:	0
DB:	10	Gaps:	0

US-09-972-758A-2 (1-359) x US-09-972-758-1 (1-1080)

```
Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360

Qy        121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 420

Qy        141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
              ||||||||||||||||||||||||||||||||||||||||||:::||||||||||||||||
Db        421 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAG 480
```

```
Qy        161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
              ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db        481 AAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAG 540

Qy        181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 600

Qy        201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 660

Qy        221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 720

Qy        241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 780

Qy        261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 840

Qy        281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArg 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 900

Qy        301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 960

Qy        321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1020

Qy        341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGAC 1077

RESULT 4
US-10-264-237-629/c
; Sequence 629, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
;  APPLICANT: Birse et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA131P1
;  CURRENT APPLICATION NUMBER: US/10/264,237
;  CURRENT FILING DATE:  2002-10-04
;  PRIOR APPLICATION NUMBER: PCT/US01/16450
;  PRIOR FILING DATE: 2001-05-18
;  PRIOR APPLICATION NUMBER: US 60/205,515
;  PRIOR FILING DATE: 2000-05-19
;  NUMBER OF SEQ ID NOS: 2876
;  SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 629
;   LENGTH: 1083
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (16)..(16)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (27)..(27)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (32)..(32)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (111)..(111)
;   OTHER INFORMATION: n equals a,t,g, or c
US-10-264-237-629

Alignment Scores:
Pred. No.:                1.06e-57        Length:          1083
Score:                    733.00          Matches:         146
Percent Similarity:       98.65%          Conservative:    0
Best Local Similarity:    98.65%          Mismatches:      2
Query Match:              38.38%          Indels:          1
DB:                       12              Gaps:            0

US-09-972-758A-2 (1-359) x US-10-264-237-629 (1-1083)

Qy        212 AspHisAspGlnGluGluProAspLeuLysThrGlyLeuTyrSerLysArgAlaAlaAla 231
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db       1081 GATCACGACCAGGAGGAGCCGGACCTCAAAACCGGCCTKTACTCCAARCGGSCCGCCGC- 1023

Qy        232 LysSerAspAspThrSerAspAspAspPheMetGluGluGlyGlyGluGluAspGlyGly 251
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1022 AAATCCGACGACACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGC 963

Qy        252 SerAspGlyMetGlyGlyAspGlySerGluPheLeuGlnArgAspPheSerGluThrTyr 271
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        962 AGCGATGGGATGGGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTAC 903

Qy        272 GluArgTyrHisThrGluSerLeuGlnAsnMetSerLysGlnGluLeuIleLysGluTyr 291
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        902 GAGCGGTACCACACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTAC 843

Qy        292 LeuGluLeuGluLysCysLeuSerArgMetGluAspGluAsnAsnArgLeuArgLeuGlu 311
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        842 CTGGAACTGGAGAAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGGAG 783

Qy        312 SerLysArgLeuGlyGlyAspAspAlaArgValArgGluLeuGluLeuGluLeuAspArg 331
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        782 AGCAAGCGGCTGGGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGG 723

Qy        332 LeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnGluLeuHisArgGlnGlnGluArg 351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        722 CTGCGCGCCGAGAACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGCGA 663

Qy        352 AlaProLeuSerLysPheGlyAsp 359
              |||||||||||||||||||||||||
Db        662 GCGCCGCTTTCCAAGTTTGGAGAC 639

RESULT 5
US-09-983-965-4895
; Sequence 4895, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 37-21(10297)C
;  CURRENT APPLICATION NUMBER: US/09/983,965
;  CURRENT FILING DATE:  2001-10-26
;  PRIOR APPLICATION NUMBER: US 09/465,231
;  PRIOR FILING DATE: 1999-12-15
;  PRIOR APPLICATION NUMBER: US 60/113,678
;  PRIOR FILING DATE: 1998-12-17
;  NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 4895
;   LENGTH: 414
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 24-LIB34-008-Q1-E1-F7
US-09-983-965-4895
```

```
Alignment Scores:
Pred. No.:                2.39e-54        Length:         414
Score:                    691.00          Matches:        130
Percent Similarity:       95.62%          Conservative:   1
Best Local Similarity:    94.89%          Mismatches:     6
Query Match:              36.18%          Indels:         0
DB:                       10              Gaps:           0

US-09-972-758A-2 (1-359) x US-09-983-965-4895 (1-414)

Qy        155 ArgArgProSerLysLysLysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGlu 174
              ||||||||||||||||||||||||  ||||||||||||||  ||||||||||||||||
Db          2 AGACGCCCCTCCAAGAAGAAGCGGCTTTGGAAACCGTACTATACGCTGACCTGGGAGGAG 61

Qy        175 LysLysLysPheAspGluLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPhe 194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 AAGAAAAAGTTCGATGAAAAACAGAGCCTGCGAGCTTCGAGGATTCGAGCCGAGATGTTC 121

Qy        195 AlaLysGlyGlnProValAlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAsp 214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 GCCAAGGGCCAGCCAGTTGCTCCCTATAACACCACGCAGTTCCTCATGGATGACCACGAC 181

Qy        215 GlnGluGluProAspLeuLysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAsp 234
              ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db        182 CAGGAGGAGCCGGATCTTAAAACCGGCCTCTATCCCAAACGGGCCGCTGCCAAATCCGAC 241

Qy        235 AspThrSerAspAspAspPheMetGluGluGlyGlyGluGluAspGlyGlySerAspGly 254
              ||||||||||||:::||||||||||||||  ||||||||||||||||||||||||||||
Db        242 GACACCAGCGATGAGGACTTTATGGAAGAAGCGGGCGAGGAGGATGGGGGCAGCGACGGG 301

Qy        255 MetGlyGlyAspGlySerGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyr 274
              ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||
Db        302 ATGGGAGGAGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGAACTATGAGCGGTAC 361

Qy        275 HisThrGluSerLeuGlnAsnMetSerLysGlnGluLeuIleLysGluTyr 291
              |||  ||||||||||||||||||||||||||||||||||||||||||||
Db        362 CACGCGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAAGAGTAC 412

RESULT 6
US-09-954-531-813
; Sequence 813, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
;  APPLICANT: Weaver, Zoe
;  TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
;  TITLE OF INVENTION:  Gene Sets
;  FILE REFERENCE: 689290-77
;  CURRENT APPLICATION NUMBER: US/09/954,531
;  CURRENT FILING DATE:  2002-05-02
;  PRIOR APPLICATION NUMBER: US/60/233,133
;  PRIOR FILING DATE: 2000-09-18
;  PRIOR APPLICATION NUMBER: US/60/234,009
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,034
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,509
;  PRIOR FILING DATE: 2000-09-22
;  PRIOR APPLICATION NUMBER: US/60/234,567
;  PRIOR FILING DATE: 2000-09-22
;  NUMBER OF SEQ ID NOS: 1392
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 813
;   LENGTH: 461
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(461)
;   OTHER INFORMATION: n=a,t,g or c
US-09-954-531-813

Alignment Scores:
Pred. No.:                1.2e-48         Length:         461
Score:                    629.00          Matches:        120
Percent Similarity:       93.85%          Conservative:   2
Best Local Similarity:    92.31%          Mismatches:     8
Query Match:              32.93%          Indels:         1
DB:                       10              Gaps:           0

US-09-972-758A-2 (1-359) x US-09-954-531-813 (1-461)

Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         72 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 131

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              ||||||||||||||||||||||||||||||||||||||||||||||||  |||||||||
Db        132 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGNCACGGAGCGGGTG 191

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 251

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 CCGGACGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 311

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        312 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 371

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
              |||||||||          :::|||||||||||||||||||||||||||||||    |||
Db        372 GC-GACTTCCCGGCNCGGCAGAAAGTGGAGCCGACGCCCGAGGCCGAGCTGCTTGNCCAG 430

Qy        121 ProCysHisAspSerGluAlaSerLysLeu 130
              |||||||||||||||||  |||||||||
Db        431 CCTTGTCATGACTCCGAGGGCAGTAAGTTG 460

RESULT 7
US-09-918-995-403/c
; Sequence 403, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 403
;   LENGTH: 495
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(495)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-403

Alignment Scores:
Pred. No.:                1.06e-47        Length:         495
Score:                    619.00          Matches:        121
Percent Similarity:       98.37%          Conservative:   0
Best Local Similarity:    98.37%          Mismatches:     2
Query Match:              32.41%          Indels:         0
DB:                       11              Gaps:           0

US-09-972-758A-2 (1-359) x US-09-918-995-403 (1-495)

Qy        237 SerAspAspAspPheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGly 256
```

```
Db        495 AGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGCATGGGGGCAGCGATGGGATGGGA 436

Qy        257 GlyAspGlySerGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThr 276

Db        435 GGGCACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACG 376

Qy        277 GluSerLeuGlnAsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLys 296

Db        375 GAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAG 316

Qy        297 CysLeuSerArgMetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGly 316

Db        315 TGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGT 256

Qy        317 GlyAspAspAlaArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsn 336

Db        255 GGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAAC 196

Qy        337 LeuGlnLeuLeuThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLys 356

Db        195 CTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAG 136

Qy        357 PheGlyAsp 359

Db        135 TTTGGAGAC 127

RESULT 8
US-10-094-749-1223
; Sequence 1223, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
;  APPLICANT: ISOGAI, TAKAO
;  APPLICANT:  SUGIYAMA, TOMOYASU
;  APPLICANT:  OTSUKI, TETSUJI
;  APPLICANT:  WAKAMATSU, AI
;  APPLICANT:  SATO, HIROYUKI
;  APPLICANT:  ISHII, SHIZUKO
;  APPLICANT:  YAMAMOTO, JUN-ICHI
;  APPLICANT:  ISONO, YUUKO
;  APPLICANT:  HIO, YURI
;  APPLICANT:  OTSUKA, KAORU
;  APPLICANT:  NAGAI, KEIICHI
;  APPLICANT:  IRIE, RYOTARO
;  APPLICANT:  TAMECHIKA, ICHIRO
;  APPLICANT:  SEKI, NAOHIKO
;  APPLICANT:  YOSHIKAWA, TSUTOMU
;  APPLICANT:  OTSUKA, MOTOYUKI
;  APPLICANT:  NAGAHARI, KENJI
;  APPLICANT:  MASUHO, YASUHIKO
;  TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
;  FILE REFERENCE: 084335/0160
;  CURRENT APPLICATION NUMBER: US/10/094,749
;  CURRENT FILING DATE:  2002-03-12
;  PRIOR APPLICATION NUMBER: 60/350,435
;  PRIOR FILING DATE: 2002-01-24
;  PRIOR APPLICATION NUMBER: JP 2001-328381
;  PRIOR FILING DATE: 2001-09-14
;  NUMBER OF SEQ ID NOS: 3381
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1223
;   LENGTH: 1330
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-094-749-1223

Alignment Scores:
Pred. No.:              1.58e-41        Length:         1330
Score:                  556.50          Matches:        141
Percent Similarity:     51.40%          Conservative:   42
Best Local Similarity:  39.61%          Mismatches:     118
Query Match:            29.14%          Indels:         55
DB:                     13              Gaps:           9

US-09-972-758A-2 (1-359) x US-10-094-749-1223 (1-1330)

Qy         20 GlyAlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArg 39

Db         87 GGAGCAGGACAGAGAAGACGGCCCCTGAATCCCATTTGGCCCCTTGCTGGCTGGAGGTGT 146

Qy         40 ValProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgPro 59

Db        147 ---------GAAAACCAGCGGTGGAGGCAGCCTTCGGGGCCTGCATTT------------ 185

Qy         60 GlyProGluGlyGluGlySerLeuGluSerGln-------------------------- 70

Db        186 ------GAGAATAAGGGTGTCACTAGTTCCAGGCGTCTGCTGAAAGATTTGGAACAGAAG 239

Qy         71 ------ProProProLeuGlnThrGlnAlaCysProGluSerSerCysLeuArgGluGly 88

Db        240 ATGATGGCCACTCCGAACCAGACCGCCTGTAATGCAGAGTCACCAGTGGCCCTGGAGGAG 299

Qy         89 GluLysGlyGlnAsnGlyAspAspSer-----------------------SerAlaGly 100

Db        300 GCCAAGACCTCTGGTGCCCCGGGGAGCCCCCAAACACCCCCTGAGCGTCATGACTCTGGT 359

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120

Db        360 GGTTCCCTGCCCCTGACACCGCGGATGGAGAGCCACTCAGAGGATGAAGATCTTGCTGGG 419

Qy        121 ProCys------------HisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGly 136

Db        420 GCTGTCGGTGGCCTGGGCTGGAACAGTAGGAGTCCCCGGACCCAGAGCCCAGGGGGCTGC 479

Qy        137 GlyGluGluGluTrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArg 156

Db        480 TCAGCGGAG-----------------GCTGTGCTGGCCCGGAAGAAACACCGTCGGCGG 521

Qy        157 ProSerLysLysLysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLys 176

Db        522 CCATCGAAGCGCAAAAGGCACTGGCGACCCTACCTGGAGCTGAGCTGGGCTGAGAAACAA 581

Qy        177 LysPheAspGluLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLys 196

Db        582 CAGCGGGATGAGAGGCAGAGCCAGAGGGCCTCCCGGGTCCGCGAAGAGATGTTCGCCAAA 641

Qy        197 GlyGlnProValAlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGlu 216

Db        642 GGCCAGCCCGTGGCCCCCTACAACACCACCCAGTTCCTGATGAATGACAGGGACCCGGAG 701

Qy        217 GluPro------AspLeuLysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAsp 234

Db        702 GAGCCCAACTTGGATGTGCCCCATGGGATCTCCCACCCAGGTTCCAGTGGGGAGAGT--- 758

Qy        235 AspThrSerAspAspAspPheMetGluGluGlyGlyGluGluAspGlyGlySerAspGly 254

Db        759 ------------------------------------GAGGCCGGGGACAGTGATGGG 779

Qy        255 MetGlyGlyAspGlySerGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyr 274

Db        780 CGGGGCCGAGCGCACGGTGAGTTCCAGCGGAAGGACTTCTCTGAGACTTACGAACGCTTC 839

Qy        275 HisThrGluSerLeuGlnAsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeu 294

Db        840 CACACCGAGAGCCTGCAGGGCCGCAGCAAGCAGGAGCTGGTGCGAGACTACCTGGAGCTG 899

Qy        295 GluLysCysLeuSerArgMetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArg 314

Db        900 GAGAAGCGGCTGTCGCAGGCGGAGGAGGAGACTAGGAGGCTGCAGCAGCTGCAGGCGTGC 959

Qy        315 LeuGlyGlyAspAspAlaArg---ValArgGluLeuGluLeuGluLeuAspArgLeuArg 333

Db        960 ACCGGCCAGCAGTCCTGCCGCCAGGTGGAGGAGCTGGCTGCCGAGGTCCAGAGGCTCCGG 1019

Qy        334 AlaGluAsnLeuGlnLeuLeuThrGluAsnGluLeuHisArgGlnGln 349
```

```
Db       1020 ACCGAAAACCAGCGGCTTCGTCAGGAGAACCAGATGTGGAACCGAGAG 1067

RESULT 9
US-10-108-260A-1496/c
; Sequence 1496, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20040005560A1el full length cDNA
;  FILE REFERENCE: H1-A0106
;  CURRENT APPLICATION NUMBER: US/10/108,260A
;  CURRENT FILING DATE:  2002-03-27
;  NUMBER OF SEQ ID NOS: 5458
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1496
;   LENGTH: 2048
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-108-260A-1496

Alignment Scores:
Pred. No.:              2.86e-40        Length:         2048
Score:                  545.00          Matches:        132
Percent Similarity:     55.77%          Conservative:   42
Best Local Similarity:  42.31%          Mismatches:     95
Query Match:            28.53%          Indels:         43
DB:                     12              Gaps:           7

US-09-972-758A-2 (1-359) x US-10-108-260A-1496 (1-2048)

Qy         54 GlnLeuGlyGlyArgProGlyProGluGlyGlu-GlySerLeuGluSerGln-------- 70
Db       1929 AAGGTAGGCTTCAGGCTAGGTAGAGAGGGAAAGGTTCCACCTCCAAGGACAGGCTGTCTG 1870

Qy         71 ----------------ProProProLeuGlnThrGlnAlaCysProGluSerSerCysLe 85
Db       1869 ATCCATCCTTCTTCCTCCTCCCTCTCTTTAGACCTCTGGTGCCCCG-------------- 1824

Qy         85 uArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGlyGlyAspPheProPr 105
Db       1823 -------GGGAGCCCCCAAACACCCCCTGAGCGTCATGACTCTGGTGGTTCCCTGCCCCT 1771

Qy        105 oProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGlnProCys-------- 122
Db       1770 GACACCGCGGATGGAGAGCCACTCAGAGGATGAAGATCTTGCTGGGGCTGTCGGTGGCCT 1711

Qy        123 ----HisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGluTr 141
Db       1710 GGGCTGGAACAGTAGGAGTCCCCGGACCCAGAGCCCAGGGGGCTGCTCAGCGGAG----- 1656

Qy        141 pGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLysLy 161
Db       1655 -------------GCTGTGCTGGCCCGGAAGAAACACCGTCGGCGGCCATCGAAGCGCAA 1609

Qy        161 sArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGluLy 181
Db       1608 AAGGCACTGGCGACCCTACCTGGAGCTGAGCTGGGCTGAGAAACAACAGCGGGATGAGAG 1549

Qy        181 sGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProValAl 201
Db       1548 GCAGAGCCAGAGGGCCTCCCGGGTCCGCGAAGAGATGTTCGCCAAAGGCCAGCCCGTGGC 1489

Qy        201 aProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluPro------As 219
Db       1488 CCCCCACAACACCACCCAGTTCCTGATGAATGACAGGGACCCGGAGGAGCCCAACTTGGA 1429

Qy        219 pLeuLysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAs 239
Db       1428 TGTGCCCCATGGGATCTCCCACCCAGGTTCCAGTGGGGAGAGT---------------- 1386

Qy        239 pAspPheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGl 259
Db       1385 ------------------------GAGGCCGGGGACAGTGATGGGCGGGGCCGAGCGCA 1351

Qy        259 ySerGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLe 279
Db       1350 CGGTGAGTTCCAGCGGAAGGACTTCTCTGAGACTTACGAACGCTTCCACACCGAGAGCCT 1291

Qy        279 uGlnAsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSe 299
Db       1290 GCAGGGCCGCAGCAAGCAGGAGCTGGTGCGAGACTACCTGGAGCTGGAGAAGCGGCTGTC 1231

Qy        299 rArgMetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAs 319
Db       1230 GCAGGCGGAGGAGGAGACTAGGAGGCTGCAGCAGCTGCAGGCGTGCACCGGCCAGCAGTC 1171

Qy        319 pAlaArg---ValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGl 338
Db       1170 CTGCCGCCAGGTGGAGGAGCTGGCTGCCGAGGTCCAGAGGCTCCGGACCGAAAACCAGCG 1111

Qy        338 nLeuLeuThrGluAsnGluLeuHisArgGlnGln 349
Db       1110 GCTTCGTCAGGAGAACCAGATGTGGAACCGAGAG 1077

RESULT 10
US-09-991-936-1712
; Sequence 1712, Application US/09991936
; Publication No. US20030073827A1
; GENERAL INFORMATION:
;  APPLICANT: Brandt, Kevin S.
;  APPLICANT:  Gaines, Patrick J.
;  APPLICANT:  Stinchcomb, Dan T.
;  APPLICANT:  Wisnewski, Nancy
;  TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES, PROTEINS AND USES THEREOF
;  FILE REFERENCE: FC-6-C1
;  CURRENT APPLICATION NUMBER: US/09/991,936
;  CURRENT FILING DATE:  2001-11-21
;  PRIOR APPLICATION NUMBER: US/09/543,668
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: 60/128,704
;  PRIOR FILING DATE: 1999-04-09
;  NUMBER OF SEQ ID NOS: 1959
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1712
;   LENGTH: 549
;   TYPE: DNA
;   ORGANISM: Ctenocephalides felis
US-09-991-936-1712

Alignment Scores:
Pred. No.:              4.91e-08        Length:         549
Score:                  184.00          Matches:        66
Percent Similarity:     42.29%          Conservative:   30
Best Local Similarity:  29.07%          Mismatches:     53
Query Match:            9.63%           Indels:         78
DB:                     11              Gaps:           12

US-09-972-758A-2 (1-359) x US-09-991-936-1712 (1-549)

Qy         34 ProGlyAlaGluGluArgValProGluGluAspSerArgTrpGlnSerArgAlaPhe--- 52
Db         33 CCAGCNGGTGAAGAAATAGTG-----------AGTAAATATCAATCAAAGTGTTTTTTA 80

Qy         53 ---ProGlnLeuGlyGlyArgProGlyProGluGlyGluGlySerLeuGluSerGlnPro 71
Db         81 TTAATACTGTTGAACATCAAAATGAGTGTAAAAATTGAA---AATATTGAATCGGAAAAT 137

Qy         72 Pro-----------------------------------------ProLeuGlnThrGlnAla 78
Db        138 CCAAAAATAATACAACCGCAGTACAGCGTCCAAAATGTCTGCCCGCTTCAAACAAGCGCT 197

Qy         79 CysProGluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSer 98
Db        198 ---------ACTGAAGCGCCTCTATCAGGAGAAAAAGGAAATGATAGTGAT--------- 239
```

```
Qy         99 AlaGlyGlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeu 118

Db        239 ------------------------------------------------------------ 239

Qy        119 AlaGlnProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGlu 138
                                 ||||||::::::::::|||      |||      |||:::
Db        240 ---------------AGAAAAGAAGCAAACCAAGTGGGT------GCAACAAATGGGAAG 278

Qy        139 GluGluTrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSer 158
                                          ||||||||||||||||
Db        279 --------------------------------AAGAAAAAACATAGACGAGGTAAATGT 305

Qy        159 LysLysLysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPhe 178
              |||:::|||   :::   |||||||||   |||      |||   ::::::|||
Db        306 AAGAGAAAACCTAAT---AAACCGTACAATAAAACAGCCTGGACCCAACGTAAGAATGTC 362

Qy        179 AspGluLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGln 198
                 :::|||:::   :::|||            ||||||::::::   |||   |||
Db        363 ---CAAAAAGAAATGATGAGAGTTCGTAGTGCCCGTGCTAAAATATTAGCCATGGGGCAC 419

Qy        199 ProValAlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluPro 218
              :::   |||   ||||||   ||||||||||||:::|||||||||
Db        420 ACATTAGTTCCTTGTAACACGAACCAATTTCTCATGGAAGATCATGAT------------ 467

Qy        219 AspLeuLysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThr------ 236
                          :::   :::|||   :::::::|||   ||||||   |||
Db        468 --------------GTCCTCACCAAGGATTCATCTGCAGACTCGGACTCTACTTAGTGT 512

Qy        237 ---SerAspAspAspPheMet 242
                 |||:::|||      |||
Db        513 CGTTCTGAAGACACTCTNATG 533

RESULT 11
US-10-084-817-158
; Sequence 158, Application US/10084817
; Publication No. US20030119009A1
; GENERAL INFORMATION:
;  APPLICANT: Susan Stuart
;  APPLICANT:  Jed G. Nuchtern
;  APPLICANT:  Sharon E. Plon
;  APPLICANT:  Jason M. Shohet
;  TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION
;  FILE REFERENCE: PA-0046 US
;  CURRENT APPLICATION NUMBER: US/10/084,817
;  CURRENT FILING DATE:  2002-02-25
;  PRIOR APPLICATION NUMBER: 60/270,784
;  PRIOR FILING DATE: 2001-02-23
;  NUMBER OF SEQ ID NOS: 365
;  SOFTWARE: PERL Program
; SEQ ID NO 158
;   LENGTH: 6354
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20030119009A1 1393778CB1
US-10-084-817-158

Alignment Scores:
Pred. No.:              7.8e-05       Length:         6354
Score:                  162.00        Matches:        99
Percent Similarity:     35.73%        Conservative:   65
Best Local Similarity:  21.57%        Mismatches:     183
Query Match:            8.48%         Indels:         113
DB:                     15            Gaps:           15

US-09-972-758A-2 (1-359) x US-10-084-817-158 (1-6354)

Qy          2 AlaGluProPheLeuSerGluTyrGln-----------------HisGlnProGln--- 14
              |||:::   :::      |||::::::                 |||   ||||||
Db       2176 GCTGATGGCTACGCTGAGGAACACGAACCCCAACTTTGTCCGCTGCATCATCCCCAACCA 2235

Qy         15 ----------------------ThrSerAsnCysThrGlyAlaAlaAlaValGlnGlu 26

                                        ::::::      |||   ||||||:::|||
Db       2236 CGAGAAGAAGGCCGGCAAGCTGGACCCGCATCTCGTGCTGGACCAGCTGCGCTGCAACGG 2295

Qy         27 -----------------GluLeuAsnProGluArgPro--------------ProGly 35
                                     |||  |||      |||              ||||||
Db       2296 TGTTCTCGAGGGCATCCGTATCTGCCGCCAGGGCTTCCCCAACAGGGTGGTCTTCCAGGA 2355

Qy         36 AlaGluGluArgValProGluGluAspSerArgTrpGlnSerArgAlaPhe------Pro 53
                     :::   :::   ||||||:::      |||:::   |||
Db       2356 GTTTCGGCAGAGATATGAGATCCTGACTCCAAACTCCATTCCCAAGGGTTTCATGGACGG 2415

Qy         54 GlnLeuGlyGlyArgProGlyProGluGlyGluGlySerLeuGluSerGlnProProPro 73
              :::   |||   |||          :::      |||:::   :::          |||
Db       2416 GAAGCAGGCGTGCGTGCTCATGATAAAAGCCCTGGAGCTCGACAGCAATCTGTACCGCAT 2475

Qy         74 LeuGlnThrGlnAlaCysProGluSerSerCys--------------------------- 84
                    |||:::            |||
Db       2476 TGGCCAGAGCAAAGTCTTCTTCCGTGCCGGTGTGCTGGCCCACCTGGAGGAGGAGCGAGA 2535

Qy         85 -----------------------------------LeuArgGluGlyGluLysGlyGln 92
                                                 |||   ::::|||      ||||||
Db       2536 CCTGAAGATCACCGACGTCATCATAGGGTTCCAGGCCTGCTGCAGGGGCTACCTGGCCAG 2595

Qy         93 AsnGlyAspAspSerSerAlaGlyGlyAspPhe---------ProProProAlaGluVal 109
                        :::|||      :::            ||||||||||||:::
Db       2596 GAAAGCATTTGCCAAGCGGCAGCAGCAGCTTACCGCCATGAAGGTCCTCCAGCGGAACTG 2655

Qy        110 GluProThrProGluAlaGluLeuLeuAla------GlnProCysHisAspSerGluAla 127
                 |||||||||      ||||||              |||      :::|||
Db       2656 CGCTGCCTACCTGAAGCTGCGGAACTGGCAGTGGTGGCGGCTCTTCACCAAGGTCAAGCC 2715

Qy        128 SerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGluTrp-----------GlyGln 143
              :::      |||   ||||||   ||||||::::::                 |||:::
Db       2716 GCTGCTGCAGGTGAGCCGGCAGGAGGAGGAGATGATGGCCAAGGAGGAGGAGCTGGTGAA 2775

Qy        144 GlnGlnArgGln---------LeuGlyLysLysLysHisArgArgArgProSerLysLys 160
                 ||||||:::          |||::::::  |||             :::
Db       2776 GGTCAGAGAGAAGCAGCTGGCTGCGGAGAACAGGCTCACGGAGATGGAGACGCTGCAGTC 2835

Qy        161 LysArgHisTrp-LysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGl 180
                    |||   :::      ||||||      :::|||:::   :::   :::
Db       2836 TCAGCTCATGGCAGAG---------AAATTGCAGCTGCAGGAGCAGCTCCAGGCAGAAAC 2886

Qy        180 uLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVa 200
               :::            |||      :::||||||   :::   ||||||   |||   ::
Db       2887 CGAGCTGTGTGCCGAGGCTGAGGAGCTCCGGGCCCGCCTGACCGCCAAGAAGCAGGAATT 2946

Qy        200 lAlaPro---------------------------------------------------- 202
               :
Db       2947 AGAAGAGATCTGCCATGACCTAGAGGCCAGGGTGGAGGAGGGGGAGGAGCGCTGCCAGCA 3006

Qy        203 ------------------------TyrAsnThrThrGlnPheLeuMetAspAspHisAs 214
                                          |||   |||   :::       ::
Db       3007 CCTGCAGGCGGAGAAGAAGAAGATGCAGCAGAACATCCAGGAGCTTGAGGAGCAGCTGGA 3066

Qy        214 pGlnGluGluProAspLeuLysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAs 234
              ::::||||||                          |||   ||||||   :::   ||
Db       3067 GGAGGAGGAG---------------AGCGCCCGGCAGAAGCT-GCAGCTGGAGAAGGTGA 3110

Qy        234 pAspThrSerAspAspAspPheMetGluGluGlyGlyGluGluAspGlyGlySerAspGl 254
               |      :::   :::          |||   :::      ||||||   :::
Db       3111 CCACCGAGGCGAAGCTGAAAAAGCTGGAGGAGGAGCAGATCATCCTGGAGGACCAGAACT 3170

Qy        254 yMetGlyGlyAspGlySerGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTy 274
                    |||   |||   |||          |||      |||      :::   :::
Db       3171 GCAAGCTGGCCAAGGAAAAGAAACTGCTGGAAGACAGAATAGCTGAGTTCACCACCAACC 3230

Qy        274 rHis-ThrGluSerLeuGlnAsnMetSerLysGlnGluLeuIleLysGluTyr------- 291
               |||    |||      ::::::   |||      :::|||      ::::::
Db       3231 TCACAGAAGAGGAGGAGAAATCTAAGAGCCTCGCCAAGCTCAAGAACAAGCATGAGGCAA 3290

Qy        292 -----LeuGluLeuGluLysCysLeuSerArgMetGluAspGluAsnAsnArgLeuArgL 310
                   :::||||||:::   |||   |||   |||   :::          |||
```

```
Db       3291 TGATCACTGACTTGGAAGAGCGCCTCCGCAGGGAGGAGAAGCAGCGACAGGAGCTGGAGA 3350

Qy        310 euGluSerLysArgLeuGlyGlyAsp--------------------------------A 319
                      ::::::|||   ||||||
Db       3351 AGACCCGCCGGAAGCTGGAGGGAGACTCCACAGACCTCAGCGACCAGATCGCCGAGCTCC 3410

Qy        319 spAlaArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnL 339
                ||||::::::   ||||||:::::::::|||   :::          |||
Db       3411 AGGCCCAGATCGCGGAGCTCAAGATGCAGCTGGCCAAGAAAGAGGAGGAGCTCCAGGCCG 3470

Qy        339 euLeuThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLys 356
                  |||          |||             |||:::          |||   |||
Db       3471 CCCTGGCCAGAGTGGAAGAGGAAGCTGCCCAGAAGAACATGGCCCTCAAGAAG 3523

RESULT 12
US-10-028-248A-35
; Sequence 35, Application US/10028248A
; Publication No. US20030235882A1
; GENERAL INFORMATION:
;  APPLICANT: Shimkets, Richard
;  APPLICANT:  Patturajan, Meera
;  APPLICANT:  Vernet, Corine
;  APPLICANT:  Casman, Stacie
;  APPLICANT:  Malyankar, Uriel
;  APPLICANT:  Shenoy, Suresh
;  APPLICANT:  Spytek, Kimberly
;  APPLICANT:  Gangolli, Esha
;  APPLICANT:  Miller, Charles
;  APPLICANT:  Boldog, Ferenc
;  APPLICANT:  Li, Li
;  APPLICANT:  Taupier Jr, Raymond J
;  APPLICANT:  Kekuda, Ramesh
;  APPLICANT:  Smithson, Glennda
;  APPLICANT:  Zerhusen, Bryan
;  APPLICANT:  Liu, Xiaohong
;  APPLICANT:  Colman, Steven
;  APPLICANT:  Tchernev, Velizar
;  APPLICANT:  Si, Jingsheng
;  APPLICANT:  Edinger, Shlomit
;  APPLICANT:  Stone, David
;  APPLICANT:  Sciore, Paul
;  APPLICANT:  Millet, Isabelle
;  APPLICANT:  Rothenberg, Mark
;  TITLE OF INVENTION: No. US20030235882A1el Nucleic Acids and Polypeptides and Methods
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 21402-222
;  CURRENT APPLICATION NUMBER: US/10/028,248A
;  CURRENT FILING DATE:  2001-12-19
;  PRIOR APPLICATION NUMBER: 60/256619
;  PRIOR FILING DATE: 2000-12-19
;  PRIOR APPLICATION NUMBER: 60/262959
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: 60/272408
;  PRIOR FILING DATE: 2001-02-28
;  PRIOR APPLICATION NUMBER: 60/285189
;  PRIOR FILING DATE: 2001-04-20
;  PRIOR APPLICATION NUMBER: 60/308039
;  PRIOR FILING DATE: 2001-07-26
;  PRIOR APPLICATION NUMBER: 60/311266
;  PRIOR FILING DATE: 2001-08-09
;  NUMBER OF SEQ ID NOS: 211
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
;   LENGTH: 7396
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-028-248A-35

Alignment Scores:
Pred. No.:              9.25e-05        Length:         7396
Score:                  162.00          Matches:        100
Percent Similarity:     36.34%          Conservative:   65
Best Local Similarity:  22.03%          Mismatches:     187

Query Match:            8.48%           Indels:         103
DB:                     12              Gaps:           15

US-09-972-758A-2 (1-359) x US-10-028-248A-35 (1-7396)

Qy          2 AlaGluProPheLeuSerGluTyrGln-----------------HisGlnProGln--- 14
              |||:::   :::       |||:::::                  |||   ||||||
Db       2107 GCTGATGGCTACGCTGAGGAACACGAACCCCAACTTTGTCCGCTGCATCATCCCCAACCA 2166

Qy         15 ------------------------ThrSerAsnCysThrGlyAlaAlaAlaValGlnGlu 26
                                      ::::::     |||   ||||||:::|||
Db       2167 CGAGAAGAAGGCCGGCAAGCTGGACCCGCATCTCGTGCTGGACCAGCTGCGCTGCAACGG 2226

Qy         27 -----------------GluLeuAsnProGluArgPro---------------ProGly 35
                               |||   |||      |||                  ||||||
Db       2227 TGTTCTCGAGGGCATCCGTATCTGCCGCCAGGGCTTCCCCAACAGGGTGGTCTTCCAGGA 2286

Qy         36 AlaGluGluArgValProGluGluAspSerArgTrpGlnSerArgAlaPhe------Pro 53
                     :::   :::   ||||||:::      |||:::   |||
Db       2287 GTTTCGGCAGAGATATGAGATCCTGACTCCAAACTCCATTCCCAAGGGTTTCATGGACGG 2346

Qy         54 GlnLeuGlyGlyArgProGlyProGluGlyGluGlySerLeuGluSerGlnProProPro 73
              :::   |||   |||          :::      |||:::   :::          |||
Db       2347 GAAGCAGGCGTGCGTGCTCATGATAAAAGCCCTGGAGCTCGACAGCAATCTGTACCGCAT 2406

Qy         74 LeuGlnThrGlnAlaCysProGluSerSerCys-------------------------- 84
                    |||:::                |||
Db       2407 TGGCCAGAGCAAAGTCTTCTTCCGTGCCGGTGTGCTGGCCCACCTGGAGGAGGAGCGAGA 2466

Qy         85 ----------------------------------LeuArgGluGlyGluLysGlyGln 92
                                                |||   :::|||      ||||||
Db       2467 CCTGAAGATCACCGACGTCATCATAGGGTTCCAGGCCTGCTGCAGGGGCTACCTGGCCAG 2526

Qy         93 AsnGlyAspAspSerSerAlaGlyGlyAspPhe---------ProProProAlaGluVal 109
                        :::|||         :::              ||||||||||||:::
Db       2527 GAAAGCATTTGCCAAGCGGCAGCAGCAGCTTACCGCCATGAAGGTCCTCCAGCGGAACTG 2586

Qy        110 GluProThrProGluAlaGluLeuLeuAla------GlnProCysHisAspSerGluAla 127
                    |||||||||      ||||||                |||      :::|||
Db       2587 CGCTGCCTACCTGAAGCTGCGGAACTGGCAGTGGTGGCGGCTCTTCACCAAGGTCAAGCC 2646

Qy        128 SerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGluTrp------------GlyGln 143
              :::      |||   ||||||   ||||||::::::                  |||:::
Db       2647 GCTGCTGCAGGTGAGCCGGCAGGAGGAGGAGATGATGGCCAAGGAGGAGGAGCTGGTGAA 2706

Qy        144 GlnGlnArgGln---------LeuGlyLysLysLysHisArgArgArgProSerLysLys 160
                 ||||||:::              |||::::::   |||              :::
Db       2707 GGTCAGAGAGAAGCAGCTGGCTGCGGAGAACAGGCTCATGGAGATGGAGACGCTGCAGTC 2766

Qy        161 LysArgHisTrp-LysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGl 180
                   |||    :::          ||||||      :::|||:::   :::   :::
Db       2767 TCAGCTCATGGCAGAG---------AAATTGCAGCTGCAGGAGCAGCTCCAGGCAGAAAC 2817

Qy        180 uLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVa 200
               :::          |||      :::||||||   :::   ||||||   |||   ::
Db       2818 CGAGCTGTGTGCCGAGGCTGAGGAGCTCCGGGCCCGCCTGACCGCCAAGAAGCAGGAATT 2877

Qy        200 lAlaPro-----------------------------------------------TyrAsnTh 205
                  :                                                  |||
Db       2878 AGAAGAGATCTGCCATGACCTAGAGGCCAGGGTGGAGGAGGAGGAGGAGCGCTACCAGCA 2937

Qy        205 rThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeuLysThrGlyLeuTy 225
               |||                  |||          :::|||:::      |||
Db       2938 CCTGCAGGCGGAGAAGAAGAAGATGCAGCAGAACATCCAGGAGCTTGAGGAGCAGCTGGA 2997

Qy        225 rSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAspPheMetGluGluGl 245
                  :::   :::|||   :::                :::   |||          |||
Db       2998 GGAGGAGGAGAGCGCCCGGCAGAAGCT-GCAGCTGGAGAAGGTGACCACCGAGGCGAAGC 3056

Qy        245 yGlyGluGluAspGlyGlySerAspGlyMetGly-----------------GlyAspGl 259
                  :::       ||||||:::|||      |||                |||   ||
Db       3057 TGAAAAAGCTGGAGGAGGAGCAGATCATCCTGGAGGACCAGAACTGCAAGCTGGCCAAGG 3116
```

```
Qy        259 ySerGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHis-ThrGluSerL 279
Db       3117 AAAAGAAACTGCTGGAAGACAGAATAGCTGAGTTCACCACCAACCTCACAGAAGAGGAGG 3176

Qy        279 euGlnAsnMetSerLysGlnGluLeuIleLysGluTyr-----------LeuGluLeuG 295
Db       3177 AGAAATCTAAGAGCCTCGCCAAGCTCAAGAACAAGCATGAGGCAATGATCACTGACTTGG 3236

Qy        295 luLysCysLeuSerArgMetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgL 315
Db       3237 AAGAGCGCCTCCGCAGGGAGGAGAAGCAGCGACAGGAGCTGGAGAAGACCCGCCGGAAGC 3296

Qy        315 euGlyGlyAsp-------------------------------AspAlaArgValArgG 324
Db       3297 TGGAGGGAGACTCCACAGACCTCAGCGACCAGATCGCCGAGCTCCAGGCCCAGATCGCGG 3356

Qy        324 luLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnG 344
Db       3357 AGCTCAAGATGCAGCTGGCCAAGAAAGAGGAGGAGCTCCAGGCCGCCCTGGCCAGAGTGG 3416

Qy        344 luLeuHisArgGlnGlnGluArgAlaProLeuSerLys 356
Db       3417 AAGAGGAAGCTGCCCAGAAGAACATGGCCCTCAAGAAG 3454
```

```
RESULT 13
US-10-107-782-35
; Sequence 35, Application US/10107782
; Publication No. US20040018970A1
; GENERAL INFORMATION:
;  APPLICANT: Boldog, Ferenc,
;  APPLICANT:  Casman, Stacie
;  APPLICANT:  Colman, Steve,
;  APPLICANT:  Edinger, Shlomit,
;  APPLICANT:  Gangolli, Esha,
;  APPLICANT:  Kekuda, Ramesh,
;  APPLICANT:  Li, Li,
;  APPLICANT:  Liu, Xiaohong,
;  APPLICANT:  Malyankar, Uriel,
;  APPLICANT:  Miller, Charles,
;  APPLICANT:  Millet, Isabelle,
;  APPLICANT:  Patturajan, Meera,
;  APPLICANT:  Rothenberg, Mark,
;  APPLICANT:  Sciore, Paul,
;  APPLICANT:  Shenoy, Suresh,
;  APPLICANT:  Shimkets, Richard,
;  APPLICANT:  Si, Jingsheng,
;  APPLICANT:  Smithson, Glennda,
;  APPLICANT:  Spytek, Kimberly,
;  APPLICANT:  Stone, David,
;  APPLICANT:  Taupier, Raymond, jr.,
;  APPLICANT:  Tchernev, Velizar,
;  APPLICANT:  Vernet, Corine,
;  APPLICANT:  Zerhusen, Brian
;  TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES AND METHODS OF USE THEREOF
;  FILE REFERENCE: 21402-222CIP
;  CURRENT APPLICATION NUMBER: US/10/107,782
;  CURRENT FILING DATE:  2002-03-27
;  PRIOR APPLICATION NUMBER: 10/028,248
;  PRIOR FILING DATE: 2001-12-19
;  PRIOR APPLICATION NUMBER: 60/256,619
;  PRIOR FILING DATE: 2000-12-19
;  PRIOR APPLICATION NUMBER: 60/262,959
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: 60/272,408
;  PRIOR FILING DATE: 2001-02-28
;  PRIOR APPLICATION NUMBER: 60/285,189
;  PRIOR FILING DATE: 2001-04-20
;  PRIOR APPLICATION NUMBER: 60/308,039
;  PRIOR FILING DATE: 2001-07-26
;  PRIOR APPLICATION NUMBER: 60/311,266
;  PRIOR FILING DATE: 2001-08-09
;  PRIOR APPLICATION NUMBER: 60/279,344
;  PRIOR FILING DATE: 2001-03-28
;  NUMBER OF SEQ ID NOS: 215
;  SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 35
;   LENGTH: 7396
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (140)..(6016)
US-10-107-782-35

Alignment Scores:
Pred. No.:                9.25e-05        Length:          7396
Score:                    162.00          Matches:         100
Percent Similarity:       36.34%          Conservative:    65
Best Local Similarity:    22.03%          Mismatches:      187
Query Match:              8.48%           Indels:          103
DB:                       12              Gaps:            15

US-09-972-758A-2 (1-359) x US-10-107-782-35 (1-7396)

Qy          2 AlaGluProPheLeuSerGluTyrGln-----------------HisGlnProGln--- 14
Db       2107 GCTGATGGCTACGCTGAGGAACACGAACCCCAACTTTGTCCGCTGCATCATCCCCAACCA 2166

Qy         15 ------------------------ThrSerAsnCysThrGlyAlaAlaAlaValGlnGlu 26
Db       2167 CGAGAAGAAGGCCGGCAAGCTGGACCCGCATCTCGTGCTGGACCAGCTGCGCTGCAACGG 2226

Qy         27 ------------------GluLeuAsnProGluArgPro---------------ProGly 35
Db       2227 TGTTCTCGAGGGCATCCGTATCTGCCGCCAGGGCTTCCCCAACAGGGTGGTCTTCCAGGA 2286

Qy         36 AlaGluGluArgValProGluGluAspSerArgTrpGlnSerArgAlaPhe------Pro 53
Db       2287 GTTTCGGCAGAGATATGAGATCCTGACTCCAAACTCCATTCCCAAGGGTTTCATGGACGG 2346

Qy         54 GlnLeuGlyGlyArgProGlyProGluGlyGluGlySerLeuGluSerGlnProProPro 73
Db       2347 GAAGCAGGCGTGCGTGCTCATGATAAAAGCCCTGGAGCTCGACAGCAATCTGTACCGCAT 2406

Qy         74 LeuGlnThrGlnAlaCysProGluSerSerCys-------------------------- 84
Db       2407 TGGCCAGAGCAAAGTCTTCTTCCGTGCCGGTGTGCTGGCCCACCTGGAGGAGGAGCGAGA 2466

Qy         85 ----------------------------------LeuArgGluGlyGluLysGlyGln 92
Db       2467 CCTGAAGATCACCGACGTCATCATAGGGTTCCAGGCCTGCTGCAGGGGCTACCTGGCCAG 2526

Qy         93 AsnGlyAspAspSerSerAlaGlyGlyAspPhe---------ProProProAlaGluVal 109
Db       2527 GAAAGCATTTGCCAAGCGGCAGCAGCAGCTTACCGCCATGAAGGTCCTCCAGCGGAACTG 2586

Qy        110 GluProThrProGluAlaGluLeuLeuAla------GlnProCysHisAspSerGluAla 127
Db       2587 CGCTGCCTACCTGAAGCTGCGGAACTGGCAGTGGTGGCGGCTCTTCACCAAGGTCAAGCC 2646

Qy        128 SerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGluTrp-----------GlyGln 143
Db       2647 GCTGCTGCAGGTGAGCCGGCAGGAGGAGGAGATGATGGCCAAGGAGGAGGAGCTGGTGAA 2706

Qy        144 GlnGlnArgGln---------LeuGlyLysLysLysHisArgArgArgProSerLysLys 160
Db       2707 GGTCAGAGAGAAGCAGCTGGCTGCGGAGAACAGGCTCATGGAGATGGAGACGCTGCAGTC 2766

Qy        161 LysArgHisTrp-LysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGl 180
Db       2767 TCAGCTCATGGCAGAG---------AAATTGCAGCTGCAGGAGCAGCTCCAGGCAGAAAC 2817

Qy        180 uLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVa 200
Db       2818 CGAGCTGTGTGCCGAGGCTGAGGAGCTCCGGGCCCGCCTGACCGCCAAGAAGCAGGAATT 2877
```

```
Qy        200 lAlaPro----------------------------------------------TyrAsnTh 205
              :                                                     |||
Db       2878 AGAAGAGATCTGCCATGACCTAGAGGCCAGGGTGGAGGAGGAGGAGGAGCGCTACCAGCA 2937

Qy        205 rThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeuLysThrGlyLeuTy 225
              |||                     |||         :::|||:::      |||
Db       2938 CCTGCAGGCGGAGAAGAAGAAGATGCAGCAGAACATCCAGGAGCTTGAGGAGCAGCTGGA 2997

Qy        225 rSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAspPheMetGluGluGl 245
              :::  :::|||  :::             :::  |||          |||
Db       2998 GGAGGAGGAGAGCGCCCGGCAGAAGCT-GCAGCTGGAGAAGGTGACCACCGAGGCGAAGC 3056

Qy        245 yGlyGluGluAspGlyGlySerAspGlyMetGly-----------------GlyAspGl 259
              :::     ||||||:::|||     |||                     |||   ||
Db       3057 TGAAAAAGCTGGAGGAGGAGCAGATCATCCTGGAGGACCAGAACTGCAAGCTGGCCAAGG 3116

Qy        259 ySerGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHis-ThrGluSerL 279
              |  |||         |||      |||      :::  :::   |||    |||
Db       3117 AAAAGAAACTGCTGGAAGACAGAATAGCTGAGTTCACCACCAACCTCACAGAAGAGGAGG 3176

Qy        279 euGlnAsnMetSerLysGlnGluLeuIleLysGluTyr------------LeuGluLeuG 295
              ::::::   |||      :::|||     ::::::                :::||||
Db       3177 AGAAATCTAAGAGCCTCGCCAAGCTCAAGAACAAGCATGAGGCAATGATCACTGACTTGG 3236

Qy        295 luLysCysLeuSerArgMetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgL 315
              ||:::  |||  |||  |||  :::         |||             ::::::|
Db       3237 AAGAGCGCCTCCGCAGGGAGGAGAAGCAGCGACAGGAGCTGGAGAAGACCCGCCGGAAGC 3296

Qy        315 euGlyGlyAsp--------------------------------AspAlaArgValArgG 324
              ||  ||||||                                |||:::::   |
Db       3297 TGGAGGGAGACTCCACAGACCTCAGCGACCAGATCGCCGAGCTCCAGGCCCAGATCGCGG 3356

Qy        324 luLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnG 344
              |||||:::::::::|||  :::         |||         |||          |
Db       3357 AGCTCAAGATGCAGCTGGCCAAGAAAGAGGAGGAGCTCCAGGCCGCCCTGGCCAGAGTGG 3416

Qy        344 luLeuHisArgGlnGlnGluArgAlaProLeuSerLys 356
              ||         |||:::       |||  |||
Db       3417 AAGAGGAAGCTGCCCAGAAGAACATGGCCCTCAAGAAG 3454

RESULT 14
US-10-264-049-543
; Sequence 543, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
;  APPLICANT: Birse et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA133P1
;  CURRENT APPLICATION NUMBER: US/10/264,049
;  CURRENT FILING DATE:  2002-10-04
;  PRIOR APPLICATION NUMBER: PCT/US01/18569
;  PRIOR FILING DATE: 2001-06-07
;  PRIOR APPLICATION NUMBER: US 60/209,467
;  PRIOR FILING DATE: 2000-06-07
;  NUMBER OF SEQ ID NOS: 4360
;  SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 543
;   LENGTH: 4041
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (162)..(162)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4022)..(4022)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4039)..(4039)
;   OTHER INFORMATION: n equals a,t,g, or c
US-10-264-049-543
```

```
Alignment Scores:
Pred. No.:              7.92e-05      Length:          4041
Score:                  159.50        Matches:         117
Percent Similarity:     32.34%        Conservative:    46
Best Local Similarity:  23.21%        Mismatches:      144
Query Match:            8.35%         Indels:          198
DB:                     12            Gaps:            20

US-09-972-758A-2 (1-359) x US-10-264-049-543 (1-4041)

Qy          3 GluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGlyAlaAla 22
              ||||||      :::|||   :::||||||   |||                  |||
Db        682 GAACCGGCAGAAGCTGAGCCTGAGCACCAAGCTCAA-----------------GCAGGT 723

Qy         23 AlaValGlnGluGluLeuAsnPro--------------------------GluArgPro 33
                :::||||||     |||                             :::
Db        724 GGAGGACGAGAAGAATTCCTTCCGGGAGCAGCTGGAGGAGGAGGAGGAGGCCAAGCACAA 783

Qy         34 ProGlyAlaGluGluArgValProGluGluAspSerArgTrpGlnSerArgAlaPhePro 53
              ||||||      :::|||   |||                                 |||
Db        784 CCTGGAGAAGCAGATCGCCACCCT---------------------------------CCA 810

Qy         54 GlnLeuGlyGlyArg---------------------------------ProGly------ 60
                ||||||||                                     ||||||
Db        811 TGCCCAGGTGGCCGACATGAAAAAGAAGATGGAGGACAGTGTGGGGTGCCTGGAAACTGC 870

Qy         61 --------------------ProGluGlyGluGlySerLeuGluSerGlnProProPro 73
                                   |||||||||   |||                |||  |||
Db        871 TGAGGAGGTGAAGAGGAAGCTCCAGAAGGACCTGGAGGG------------CCTGAGCCA 918

Qy         74 LeuGlnThrGlnAlaCysProGluSerSerCysLeuArgGluGlyGluLysGlyGlnAsn 93
                  |||       |||         ||||||:::      :::   |||:::
Db        919 ---------GCGGCACGAGGAGAAGGTGGCCGCCTACGACAAGCTGGAGAAGACCAAGAC 969

Qy         94 ------------GlyAspAspSerSerAlaGlyGlyAspPheProPro---ProAlaGlu 108
                         |||              |||||||||      ||||||   |||
Db        970 GCGGCTGCAGCAGGAGCTGGACGACCTGCTGGTGGACCTGGACCACCAGCGCCAGAGCGC 1029

Qy        109 ValGluProThrProGluAlaGlu---------------------------------Leu 117
              |||:::|||      |||||||||
Db       1030 GTGCAACCTGGAGAAGAAGCAGAAGAAGTTTGACCAGCTCCTGGCGGAGGAGAAGACCAT 1089

Qy        118 LeuAlaGlnProCysHisAspSerGluAlaSerLysLeuGlyAlaPro------------ 133
              |||   |||  |||      :::                 |||   |||
Db       1090 CTCTGCCAAGTATGCAGAGGAGCGCGACCGGGCTGAGGCGGAGGCCCGAGAGAAGGAGAC 1149

Qy        134 ----------------------------------------AlaAlaGlyGlyGluGlu 139
                                                     |||   |||||  |||
Db       1150 CAAGGCTCTGTCGCTGGCCCGGGCCCTGGAGGAAGCCATGGAGCAGAAGGCGGACTGGAG 1209

Qy        140 GluTrpGlyGlnGlnGlnArg---------------------------------GlnLeu 148
                 :::|||   |||                                    :::
Db       1210 CGGCTCAACAAGCAGTTCCGCACGGAGATGGAGGACCTTATGAGCTCCAAGGATGATGTG 1269

Qy        149 GlyLysLysLysHisArgArgArgProSerLysLysLysArgHisTrpLysProTyrTyr 168
              ||||||     |||          ||||||:::          :::
Db       1270 GGCAAGAGTGTCCACGAGCTGGAGAAGTCCAAGCGGGCCCTAGAGCAGCAGGTGGAGGAG 1329

Qy        169 LysLeuThrTrpGluGluLysLysLysPheAspGluLysGlnSerLeuArgAlaSerArg 188
                 |||      |||:::   :::   ||||||   |||:::        ::::::
Db       1330 ATGAAGACGCAGCTGGAAGAGCTGGAG---GACGAGCTGCAGGCCACCGAAGATGCCAAG 1386

Qy        189 IleArgAlaGluMetPheAlaLysGlyGlnProValAlaProTyrAsnThrThrGlnPhe 208
              :::|||   |||:::      :::
Db       1387 CTGCGGTTGGAGGTCAACCTGCAGGCCATGAAGGCC----------------------- 1422

Qy        209 LeuMetAspAspHisAspGlnGluGluProAspLeu-LysThrGlyLeuTyrSerLysAr 228
                          |||   |||   |||||| :::   |||             ||
Db       1423 -----------------CAGTTCGAGCGGGACCTGCAGGGCCGGGACGAGCAGAGCGAG 1464

Qy        228 gAlaAlaAlaLysSer-AspAspThrSerAspAspAspPheMetGluGluGlyGlyGluG 248
              |             |||       |||         |||            ||||||   |
```

```
Db       1465 GAGAAGAAGAAGCAGCTGGTCAGACAGGTGCGGGAGATGGAGGCAGAGCTGGAGGACGAG 1524

Qy        248 luAspGlyGlySerAspGlyMetGly---------------GlyAspGlySerGlu-Phe 262
              ||:::           ||||||  |||              ||||||||   |||
Db       1525 AGGAAGCAGCGCTCGATGGCAGTGGCCGCCCGGAAGAAGCTGGAGATGGACCTGAAGGAC 1584

Qy        263 LeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGlnAsnMet 282
              |||:::           ::::::      :::         |||:::::::::   :::
Db       1585 CTGGAGGCGCACATC---GACTCGGCCAACAAGAACCGGGACGAAGCCATCAAACAGCTG 1641

Qy        283 ---------------------------------------------------------Ser 283
                                                                       |||
Db       1642 CGGAAGCTGCAGGCCCAGATGAAGGACTGCATGCGCGAGCTGGATGACACCCGCGCCTCT 1701

Qy        284 LysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArgMetGluAsp 303
              ::::::|||::::::   :::         |||   ||||||   |||      ||||||
Db       1702 CGTGAGGAGATCCTGGCCCAGGCCAAAGAGAACGAGAAGAAGCTGAAGAGCATGGAGGCC 1761

Qy        304 GluAsnAsnArgLeuArgLeuGluSer--------------------------------- 312
              |||       :::|||:::   |||
Db       1762 GAGATGATCCAGTTGCAGGAGGAACTGGCAGCCGCGGAGCTGCCAAGCGCCAGGCCCAGC 1821

Qy        312 ------------------------------------------------------------ 312

Db       1822 AGGAGCGGGATGAGCTGGCTGACGAGATCGCCAACAGCAGCGGCAAAGGAGCCCTGGCGT 1881

Qy        313 ---LysArgLeuGlyGlyAspAspAlaArgValArgGluLeuGluLeuGluLeu------ 329
                 :::|||   |||      :::|||||:::     ||||||   ||||||
Db       1882 TAGAGGAGAAGCGGCGTCTG-GAGGCCCGCATCGCCCAGCTGGAGGAGGAGCTGGAGGAG 1940

Qy        330 --------------------------AspArgLeuArgAlaGluAsnLeuGlnLeu--- 339
                                        |||||||||:::     |||||||||:::
Db       1941 GAGCAGGGCAACACGGAGCTGATCAACGACCGGCTGAAGAAGGCCAACCTGCAGATCGAC 2000

Qy        340 -----------------------------------LeuThrGluAsnGluLeuHisArg 347
                                                 :::||||||      |||
Db       2001 CAGATCAACACCGACCTGAACCTGGAGCGCAGCCACGCCCAGAAGAACGAGAATGCTCGG 2060

Qy        348 GlnGln 349
              ||||||
Db       2061 CAGCAG 2066
```

RESULT 15
US-10-085-117-321
; Sequence 321, Application US/10085117
; Publication No. US20030232334A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 321
; LENGTH: 2901
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-085-117-321

Alignment Scores:

Pred. No.:	0.00014	Length:	2901
Score:	155.00	Matches:	105
Percent Similarity:	39.39%	Conservative:	51
Best Local Similarity:	26.52%	Mismatches:	157
Query Match:	8.12%	Indels:	83
DB:	12	Gaps:	23

US-09-972-758A-2 (1-359) x US-10-085-117-321 (1-2901)

```
Qy          2 AlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGlyAla 21
              :::||||||      :::|||   |||   :::                  ||||||
Db       1728 TCGGAACCAGACGCAGCTGAGCAGCAAGCACGAGCT------------GCAGCTGGAGCA 1775

Qy         22 AlaAlaVal-----------------------GlnGluGluLeu---------AsnPro 30
                 |||                         |||||||||:::
Db       1776 GATGCACAAACGTTTTGAACAAGAAATCAACGCCAAGAAGAAATTCTATGACGTGGAGCT 1835

Qy         31 GluArgProProGlyAlaGluGluArgValProGluGluAspSerArgTrpGlnSerArg 50
                 :::|||      |||   |||            ||||||               |||
Db       1836 AGAAAACCTGGAGCGGCAGCAGAAGCAGCAGGTAGAGAAGATGGAGCAGGACCACAGCGT 1895

Qy         51 AlaPheProGlnLeuGlyGlyArgProGlyProGluGlyGluGlySerLeuGluSerGln 70
              |||   ||||||.  ||||||:::         |||   |||   ||||||      |||
Db       1896 GCG---CCGCAAAGAGGAGGCCAAGCGGATCCGCCTGGAGCAGGATCGAGACTACGCCAA 1952

Qy         71 ProProProLeuGlnThrGlnAlaCysProGluSerSerCysLeuArgGluGlyGluLys 90
                 |||      |||||||||      |||                  ||||||:::
Db       1953 ---GTTCCAAGAGCAACTCAAGCAAATGAAGAAGGA---------------GGTGAAGAG 1994

Qy         91 ---GlyGlnAsnGlyAspAspSerSerAlaGlyGlyAspPheProProProAlaGluVal 109
                 |||         ::::::|||   |||:::               ||||||
Db       1995 TGAGGTTGAGAAACTGCCCCGGCAGCAGCGGAAGGAGAGCAT------GAAGCAGAAGAT 2048

Qy        110 GluProThrProGluAlaGluLeuLeuAlaGlnPro------CysHisAspSerGluAla 127
                 |||      |||      |||   |||         |||      :::|||
Db       2049 GGAGGAACACTCTCAGAAGAAACAGCGGCTGGACCGGGACTTTGTAGCCAAGCAGAA--- 2105

Qy        128 SerLysLeuGlyAlaProAlaAlaGlyGlyGluGlu---GluTrpGlyGlnGlnGlnArg 146
                    |||   |||   ||||||   ||||||         |||:::|||   |||
Db       2106 ---------GGAGGACCTGGAGCTGGCCATGAGGAAGCTCACCACGGAAAACAGGCGCGA 2156

Qy        147 GlnLeuGlyLysLysLysHisArgArgArgPro-SerLysLysLysArgHisTrpLysPr 166
              |||         :::      |||   |||   |||||||||:::            :::
Db       2157 GATCTG------TGACAAGGAGCGAGATTGCCTTAGCAAGAAGCAGGAGCTCCTCCGAGA 2210

Qy        166 oTyrTyrLysLeuThrTrpGlu---GluLysLysLysPheAspGluLysGlnSerLeuAr 185
                              ||||||   |||:::      |||:::         |||
Db       2211 CCGTGAGGCGGCCCTGTGGGAGATGGAGGAGCACCAGTTGCAGGAGCGACATCAGCTGGT 2270

Qy        185 gAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProValAlaProTyrAsnTh 205
                   :::::::::   :::   |||   :::
Db       2271 GAAGCAGCAACTGAAGGACCAGTACTTCCTGCAGCGGCATGAC----------------- 2313

Qy        205 rThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeuLysThrGlyLeuTy 225
                         |||:::      |||::::::|||         ::::::         ||
Db       2314 ----------CTGCTGCGTAAGCATGAGAAGGAGCGGGAACAGATGCAGCGC------TA 2357

Qy        225 rSerLysArg------AlaAlaAlaLysSerAspAspThrSerAspAspAspPheMetGl 243
              |::::::|||              |||                :::         :::
Db       2358 CAACCAGCGCATGATGGAGCAGCTGAAGGTCAGACAGCAACAGGAGAAGGCGCGGCTACC 2417

Qy        243 uGluGlyGlyGluGluAspGly-----------------------GlySerAspGlyMe 255
               :::            ||||||                       |||      ::
Db       2418 CAAGATCCAGAGGAGCGACGGCGAGACGCGCATGGCCATGTACAAGAAGAGCCTGCACAT 2477

Qy        255 tGlyGlyAspGlySerGluPheLeuGlnArgAsp------------PheSerGluThrTy 271
              :  |||   ||||||         ||||||:::            ||||||:::
Db       2478 CAACGGTGCGGGCAGTGCCTCTGAGCAGCGCGAGAAGATCAAGCAGTTCTCCCAGCAGGA 2537

Qy        271 rGlu---ArgTyrHisThrGluSerLeuGlnAsnMetSerLysGlnGlu----------- 286
               |||   |||         |||   ||||||         |||   |||
Db       2538 GGAGAAGAGGCAGAAGGCGGAGAGGCTGCAGCAGCAGCAGAAACACGAGCACCAGATGCG 2597

Qy        287 ----LeuIleLys--------------------------GluTyrLeuGluLeuGluLy 296
                  ::::::                             :::   |||:::   |||||
Db       2598 AGACATGGTGGCGCAGTGTGAAAGCAACATGAGCGAGCTACAGCAGCTGCAGAATGAAAA 2657

Qy        296 sCysLeuSerArgMetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGl 316
              ||||            :::|||   |||      :::|||:::
Db       2658 GTGTTATCTGTTGGTAGAGCATGAAACCCAGAAGCTGAAG------------------GC 2699

Qy        316 yGlyAspAspAlaArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAs 336
```

```
                |||::::::      :::   |||:::         |||:::||||||
Db       2700 CCTGGATGAGAGCCACAACCAGAGCCTGAAGGAATGGCGAGACAAGCTTCGGCCACGCAA 2759

Qy        336 nLeuGlnLeuLeuThrGlu---AsnGluLeuHisArgGlnGlnGlu 350
                 |||       :::   |||:::       |||:::||||||
Db       2760 AAAGGCCCTGGAAGAGGATTTGAACCAGAAGAAGCGGGAACAGGAA 2805


Search completed: February  5, 2004, 17:11:50
Job time : 507 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 5, 2004, 13:39:35 ; Search time 94 Seconds
(without alignments)
1685.708 Million cell updates/sec

Title: US-09-972-758A-2
Perfect score: 1910
Sequence: 1 MAEPFLSEYQHQPQTSNCTG.....LTENELHRQQERAPLSKFGD 359

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09972758/runat_05022004_095004_1219/app_query.fasta_1.519
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09972758_@CGN_1_1_56_@runat_05022004_095004_1219 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	158.5	8.3	9551	1	US-08-056-200-93	Sequence 93, Appl
2	158.5	8.3	9551	2	US-08-800-644-93	Sequence 93, Appl
3	155	8.1	2277	1	US-08-676-967-5	Sequence 5, Appli
4	155	8.1	2277	1	US-08-676-974-5	Sequence 5, Appli
5	155	8.1	2277	2	US-09-098-487-5	Sequence 5, Appli
6	147	7.7	7453	4	US-09-620-312D-248	Sequence 248, App
7	147	7.7	7501	4	US-09-620-312D-249	Sequence 249, App
8	142	7.4	2820	4	US-09-252-991A-2538	Sequence 2538, Ap
9	141	7.4	2304	1	US-08-464-266-1	Sequence 1, Appli
10	141	7.4	2304	1	US-08-464-272-1	Sequence 1, Appli
11	141	7.4	2304	3	US-08-464-514-1	Sequence 1, Appli
12	141	7.4	2304	3	US-08-486-403-1	Sequence 1, Appli
13	137	7.2	1617	4	US-09-265-013-2	Sequence 2, Appli
14	137	7.2	7308	3	US-09-011-745-3	Sequence 3, Appli
15	137	7.2	7308	3	US-09-011-745-4	Sequence 4, Appli
16	137	7.2	7616	3	US-09-011-745-2	Sequence 2, Appli
c 17	137	7.2	8202	1	US-08-258-420-13	Sequence 13, Appl
18	137	7.2	8332	3	US-08-850-961-1	Sequence 1, Appli
19	137	7.2	8332	4	US-09-479-776-1	Sequence 1, Appli
20	137	7.2	8332	4	US-09-309-572-11	Sequence 11, Appl
21	137	7.2	8332	4	US-09-315-127-1	Sequence 1, Appli
22	137	7.2	8332	4	US-09-265-013-1	Sequence 1, Appli
23	137	7.2	8332	4	US-09-554-572-25	Sequence 25, Appl
24	134	7.0	3930	3	US-09-162-373-2	Sequence 2, Appli
25	134	7.0	3930	3	US-09-467-946-2	Sequence 2, Appli
26	133.5	7.0	6755	3	US-08-931-999-4	Sequence 4, Appli
27	131.5	6.9	5661	4	US-08-938-105-2	Sequence 2, Appli
28	131	6.9	3023	4	US-09-593-589-10	Sequence 10, Appl
29	131	6.9	6775	4	US-09-620-312D-289	Sequence 289, App
30	130	6.8	1835	4	US-09-216-393B-80	Sequence 80, Appl
31	128.5	6.7	2581	4	US-09-370-838-66	Sequence 66, Appl
c 32	127	6.6	2139	4	US-09-252-991A-2295	Sequence 2295, Ap
33	126	6.6	7791	3	US-08-949-386-23	Sequence 23, Appl
34	126	6.6	7791	3	US-08-450-562-23	Sequence 23, Appl
35	126	6.6	7791	4	US-08-984-709A-23	Sequence 23, Appl
36	126	6.6	7791	4	US-08-450-272-23	Sequence 23, Appl
37	126	6.6	7808	2	US-08-149-097D-22	Sequence 22, Appl
38	126	6.6	7808	3	US-08-949-386-22	Sequence 22, Appl
39	126	6.6	7808	3	US-08-450-562-22	Sequence 22, Appl
40	126	6.6	7808	4	US-08-984-709A-22	Sequence 22, Appl
41	126	6.6	7808	4	US-08-450-272-22	Sequence 22, Appl
42	125.5	6.6	2384	1	US-07-814-964-10	Sequence 10, Appl
43	125.5	6.6	2384	1	US-08-258-442-10	Sequence 10, Appl
44	125.5	6.6	2384	1	US-08-328-809-5	Sequence 5, Appli
45	125.5	6.6	2384	4	US-08-866-840-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-08-056-200-93
; Sequence 93, Application US/08056200
; Patent No. 5616500
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; TITLE OF INVENTION: Mehods of Using Same
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/056,200
; FILING DATE: 30-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fedrick, Michael F.
; REGISTRATION NUMBER: 36,799
; REFERENCE/DOCKET NUMBER: NIH054.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502

```
;  INFORMATION FOR SEQ ID NO:  93:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9551 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1507..1644
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  1645..2511
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2512..8070
US-08-056-200-93

Alignment Scores:
Pred. No.:                0.002          Length:          9551
Score:                    158.50         Matches:         91
Percent Similarity:       39.89%         Conservative:    51
Best Local Similarity:    25.56%         Mismatches:      135
Query Match:              8.30%          Indels:          80
DB:                       1              Gaps:            16

US-09-972-758A-2 (1-359) x US-08-056-200-93 (1-9551)

Qy         31 GluArgProProGlyAlaGluGluArg----------ValProGluGlu 43
              ||||||        ||||||||||||          :::|||||||||
Db       4810 GAGAGGCAGCTGAGGGCCGAGGAGCGCCAGCAGCGGGAACAACGGTTTCTCCCGGAGGAG 4869

Qy         44 AspSerArgTrpGlnSerArgAlaProGlnLeuGlyGlyArgProGlyProGluGly 63
              :::   :::   |||                        ||||||         |||
Db       4870 GAGGAGAAGGAGCAGCGC--------------------GGCCGCCAGCGACGCGAGAGG 4908

Qy         64 GluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysProGluSerSer 83
              |||      |||:::                  |||   |||           :::
Db       4909 GAGAAAGAGCTGCAGTTCCTGGAGGAAGAGGAGCAGCTCCAGCGGCGGGAGCGTGCCCAA 4968

Qy         84 CysLeuArgGluGlyGluLysGly---GlnAsnGlyAspAspSerSerAlaGlyGlyAsp 102
                 |||:::|||   |||   |||   |||   |||         |||   |||
Db       4969 CAGCTCCAGGAGGAGGAGGACGGCCTCCAGGA-GGATCAGGAGAGGAGGCGACAGGAGCA 5027

Qy        103 PhePro---ProProAlaGluVal--------------------GluProThrProGlu 114
              |||   |||       ||||||                     |||       |||
Db       5028 GCGCCGCGACCAAAAATGGAGGTGGCAACTAGAAGAAGAAAGGAAGAGACGCCGCCACAC 5087

Qy        115 AlaGluLeuLeuAlaGlnProCysHisAspSerGluAlaSerLysLeuGlyAlaProAla 134
              |||          |||   |||        :::|||                |||   |||
Db       5088 GCTGTACGCCAAGCCAGCCCTACAAGAGCAGCTGAGGAAGGAACA------GCAGCTGCT 5141

Qy        135 AlaGlyGly-GluGluGluTrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisAr 154
              ||||||||| |||||||||      :::::::::|||:::                  ::
Db       5142 GCAGGAGGAGGAGGAGGAGCTACAGAGAGAGGAGCGCGAG-----------------AA 5183

Qy        154 gArgArgProSerLysLysLysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGl 174
              :||||||       :::::::::|||   ::::::         :::|||      |||||
Db       5184 GAGAAGGCGCCAAGAACAGGAGAGACAATACCGCGAGGAAGAGCAGCTGCAGCAGGAGGA 5243

Qy        174 uLysLysLysPheAspGluLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPh 194
              |:::::       |||::::::      |||      :::   |||          ::
Db       5244 AGAGCAGCTGCTGAGAGAGGAACGGGAGAAAAGAAGACGCCAGGAGCGGGAAAGGCAATA 5303

Qy        194 eAlaLysGlyGlnProValAlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAs 214
              :   |||   :::                              |||          ::
Db       5304 TCGGAAGGATAAGAAG--------------------------CTGCAGCAGAAGGAAGA 5336

Qy        214 pGln---------GluGluProAspLeuLysThrGlyLeuTyrSerLysArgAlaAlaAl 231
              :|||         |||||||||:::   :::               :::|||
Db       5337 GCAGCTGCTGGGAGAGGAACCGGAGAAGAGA--------------AGGCGCCAGGAGCG 5381

Qy        231 aLysSerAspAspThrSerAspAspAspPheMetGluGluGlyGlyGluGluAspGlyGl 251
               :::                  :::::::::      :::|||
Db       5382 GGAGAAAAAATACCGCGAGGAAGAGGAGTTGCAGCAGGAG------------------- 5421

Qy        251 ySerAspGlyMetGlyGlyAspGlySerGluPheLeuGlnArgAspPheSerGluThrTy 271
                                          |||   |||   |||:::      |||
Db       5422 ------------------------GAAGAGCAGCTGCTGAGAGAGGAACGGGAGAAGAG 5456

Qy        271 rGluArg-----------------TyrHisThrGluSerLeuGlnAsnMetSerLysGl 285
               |||                  |||      :::             :::::
Db       5457 AAGGCGCCAGGAGTGGGAGAGGCAGTACCGCAAAAAAGACGAGCTGCAGCAGGAAGAAGA 5516

Qy        285 nGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArgMetGluAsp----- 303
              ::::|||::::::|||      |||      :::         |||
Db       5517 GCAGCTGCTGAGAGAGGAACGGGAGAAAAGAAGACTCCAGGAGCGGGAGAGGCAATATCG 5576

Qy        304 -GluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAlaArgValAr 323
               |||         |||:::   |||   ::::::|||   |||::::::          ||
Db       5577 GGAGGAAGAGGAGCTGCAGCAGGAGGAAGAGCAGCTGCTGGGAGAGGAACGGGAGACGAG 5636

Qy        323 g---------GluLeuGluLeuGluLeu---------AspArgLeuArgAlaGluAsnLe 337
              |         |||||||||   :::             :::   |||:::   |||
Db       5637 AAGGCGCCAGGAGCTGGAGAGGCAATATCGGAAGGAAGAGGAGCTGCAGCAGGAGGAAGA 5696

Qy        337 uGlnLeuLeuThrGluAsn---GluLeuHisArgGlnGlnGluArg 351
              |||||||||   |||      |||      |||:::|||||||||
Db       5697 GCAGCTGCTGAGAGAGGAACCGGAGAAGAGAAGGCGCCAGGAGCGG 5742

RESULT 2
US-08-800-644-93
; Sequence 93, Application US/08800644
; Patent No. 5958752
;  GENERAL INFORMATION:
;    APPLICANT:  Steinert, Peter M.
;    APPLICANT:  Lee, Seung-Chul
;    APPLICANT:  Kim, In-Gyu
;    APPLICANT:  Chung, Soo-Il
;    APPLICANT:  Park, Sang-Chul
;    TITLE OF INVENTION:  Trichohyalin and Transglutaminase-3 and
;    TITLE OF INVENTION:  Mehods of Using Same
;    NUMBER OF SEQUENCES:  117
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  620 Newport Center Drive, Sixteenth Floor
;      CITY:  Newport Beach
;      STATE:  CA
;      COUNTRY:  U.S.A.
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/800,644
;      FILING DATE:  14-FEB-1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/056,200
;      FILING DATE:  30-APR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fedrick, Michael F.
;      REGISTRATION NUMBER:  36,799
;      REFERENCE/DOCKET NUMBER:  NIH054.001A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (714) 760-0404
;      TELEFAX:  (714) 760-9502
;  INFORMATION FOR SEQ ID NO:  93:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9551 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1507..1644
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  1645..2511
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2512..8070
US-08-800-644-93

Alignment Scores:
Pred. No.:              0.002          Length:         9551
Score:                  158.50         Matches:        91
Percent Similarity:     39.89%         Conservative:   51
Best Local Similarity:  25.56%         Mismatches:     135
Query Match:            8.30%          Indels:         80
DB:                     2              Gaps:           16

US-09-972-758A-2 (1-359) x US-08-800-644-93 (1-9551)

Qy         31 GluArgProProGlyAlaGluGluArg----------------------ValProGluGlu 43
              ||||||       ||||||||||||                        :::||||||||||
Db       4810 GAGAGGCAGCTGAGGGCCGAGGAGCGCCAGCAGCGGGAACAACGGTTTCTCCCGGAGGAG 4869

Qy         44 AspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGlyProGluGly 63
              :::   :::   |||                     ||||||          |||
Db       4870 GAGGAGAAGGAGCAGCGC---------------------GGCCGCCAGCGACGCGAGAGG 4908

Qy         64 GluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysProGluSerSer 83
              |||      |||:::                    |||   |||           :::
Db       4909 GAGAAAGAGCTGCAGTTCCTGGAGGAAGAGGAGCAGCTCCAGCGGCGGGAGCGTGCCCAA 4968

Qy         84 CysLeuArgGluGlyGluLysGly---GlnAsnGlyAspAspSerSerAlaGlyGlyAsp 102
                |||:::|||   |||   |||   |||   |||             |||   |||
Db       4969 CAGCTCCAGGAGGAGGAGGACGGCCTCCAGGA-GGATCAGGAGAGGAGGCGACAGGAGCA 5027

Qy        103 PhePro---ProProAlaGluVal--------------------GluProThrProGlu 114
                |||   |||      ||||||                     |||      |||
Db       5028 GCGCCGCGACCAAAAATGGAGGTGGCAACTAGAAGAAGAAAGGAAGAGACGCCGCCACAC 5087

Qy        115 AlaGluLeuLeuAlaGlnProCysHisAspSerGluAlaSerLysLeuGlyAlaProAla 134
              |||          |||   |||         :::|||                |||   |||
Db       5088 GCTGTACGCCAAGCCAGCCCTACAAGAGCAGCTGAGGAAGGAACA------GCAGCTGCT 5141

Qy        135 AlaGlyGly-GluGluGluTrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisAr 154
              |||||||| |||||||||      :::::::::|||:::                   ::
Db       5142 GCAGGAGGAGGAGGAGGAGCTACAGAGAGAGGAGCGCGAG-----------------AA 5183

Qy        154 gArgArgProSerLysLysLysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGl 174
              :||||||       :::::::::|||   ::::::         :::|||       |||||
Db       5184 GAGAAGGCGCCAAGAACAGGAGAGACAATACCGCGAGGAAGAGCAGCTGCAGCAGGAGGA 5243

Qy        174 uLysLysLysPheAspGluLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPh 194
              |::::::         |||::::::      |||      :::   |||          ::
Db       5244 AGAGCAGCTGCTGAGAGAGGAACGGGAGAAAAGAAGACGCCAGGAGCGGGAAAGGCAATA 5303

Qy        194 eAlaLysGlyGlnProValAlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAs 214
              :   |||   :::                          |||                ::
Db       5304 TCGGAAGGATAAGAAG-------------------------CTGCAGCAGAAGGAAGA 5336

Qy        214 pGln---------GluGluProAspLeuLysThrGlyLeuTyrSerLysArgAlaAlaAl 231
              :|||         |||||||||:::   :::               :::|||
Db       5337 GCAGCTGCTGGGAGAGGAACCGGAGAAGAGA---------------AGGCGCCAGGAGCG 5381

Qy        231 aLysSerAspAspThrSerAspAspAspPheMetGluGluGlyGlyGluGluAspGlyGl 251
               :::               :::::::::      :::|||
Db       5382 GGAGAAAAAATACCGCGAGGAAGAGGAGTTGCAGCAGGAG------------------- 5421

Qy        251 ySerAspGlyMetGlyGlyAspGlySerGluPheLeuGlnArgAspPheSerGluThrTy 271
                                              |||   |||   |||:::      |||
Db       5422 ------------------------GAAGAGCAGCTGCTGAGAGAGGAACGGGAGAAGAG 5456

Qy        271 rGluArg-----------------TyrHisThrGluSerLeuGlnAsnMetSerLysGl 285
               |||                  |||      :::                    :::::
Db       5457 AAGGCGCCAGGAGTGGGAGAGGCAGTACCGCAAAAAAGACGAGCTGCAGCAGGAAGAAGA 5516

Qy        285 nGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArgMetGluAsp----- 303
              ::::|||::::::|||      |||      :::          |||
Db       5517 GCAGCTGCTGAGAGAGGAACGGGAGAAAAGAAGACTCCAGGAGCGGGAGAGGCAATATCG 5576

Qy        304 -GluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAlaArgValAr 323
               |||         |||:::   |||   ::::::|||   |||::::::          ||
Db       5577 GGAGGAAGAGGAGCTGCAGCAGGAGGAAGAGCAGCTGCTGGGAGAGGAACGGGAGACGAG 5636

Qy        323 g---------GluLeuGluLeuGluLeu---------AspArgLeuArgAlaGluAsnLe 337
              |         |||||||||   :::             :::   |||:::   |||
Db       5637 AAGGCGCCAGGAGCTGGAGAGGCAATATCGGAAGGAAGAGGAGCTGCAGCAGGAGGAAGA 5696

Qy        337 uGlnLeuLeuThrGluAsn---GluLeuHisArgGlnGlnGluArg 351
               |||||||||   |||      |||      |||:::|||||||||
Db       5697 GCAGCTGCTGAGAGAGGAACCGGAGAAGAGAAGGCGCCAGGAGCGG 5742

RESULT 3
US-08-676-967-5
; Sequence 5, Application US/08676967
; Patent No. 5747317
;  GENERAL INFORMATION:
;    APPLICANT:  COLLINS, KATHLEEN
;    TITLE OF INVENTION:  Human Telomerase
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Science & Technology Law Group
;      STREET:  268 Bush Street, Suite 3200
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/676,967
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman Ph.D., Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  UCB96-055
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)343-4341
;      TELEFAX:  (415)343-4342
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2277 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-676-967-5

Alignment Scores:
Pred. No.:              0.00055        Length:         2277
Score:                  155.00         Matches:        96
Percent Similarity:     33.93%         Conservative:   55
Best Local Similarity:  21.57%         Mismatches:     147
Query Match:            8.12%          Indels:         148
DB:                     1              Gaps:           18

US-09-972-758A-2 (1-359) x US-08-676-967-5 (1-2277)
```

```
Qy          6 LeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGlyAlaAlaAlaValGln 25
              |||   :::  ||||||                          |||      |||   :::
Db        144 CTACGTGACCTTCAGCAT--------------------GCTGGAGGACGTGCAGCGCGC 182

Qy         26 GluGlu--------------------LeuAsnProGluArgProProGlyAlaGluGlu 38
                |||                  |||     :::|||      |||  ||||||
Db        183 CCTGAAGGAGATCACCACCTTCGAGGGCTGCAAGATCAACGTGACCGTGGCCAAGAAGAA 242

Qy         39 ArgValProGluGluAspSerArgTrpGlnSerArg------AlaProGlnLeuGly 56
                 ::::::              |||  |||           ||||||  |||
Db        243 GCTGCGCAACAAGACCAAGGAGAAGGGCAAGAACGAGAACAGCGAGTGCCCCAAGAAGGA 302

Qy         57 GlyArgProGlyProGluGlyGluGlySerLeuGluSerGlnPro--------------- 71
                :::  |||  ||||||:::|||     :::     |||
Db        303 GCCCAA---GGCCAAGAAGGCCAAGGTGGCCGACAAGAAGGCCCGCCTGATCATCCGCAA 359

Qy         72 ProProLeuGlnThrGlnAlaCysProGluSerSerCysLeuArgGluGlyGluLysGly 91
              |||  ||||||  |||              :::|||       :::
Db        360 CCTGAGCTTCAAGTGCAGCGAGGACGACCTGAAGACCGTGTTCGCCCAGTTCGGCGC--- 416

Qy         92 GlnAsnGlyAspAspSerSerAlaGlyGlyAspPheProProProAlaGluValGluPro 111
                          |||||||||:::  ||||||  |||     :::
Db        417 -----------------CGTGCTGGAGGTGAACATCCCCCGCAAGCCCGACGGCAA--- 455

Qy        112 ThrProGluAlaGluLeuLeuAlaGlnProCysHisAspSerGluAlaSerLysLeuGly 131
               :::|||   |||      :::         :::            :::   |||
Db        456 ------GATGCGCGGCTTCGGCTTCGTGCAGTTCAAGAACCTGCTGGAGGCCGGCAAGGC 509

Qy        132 AlaProAlaAlaGlyGlyGluGluGlu------------------------------- 140
                       |||  |||  |||
Db        510 ------CCTGAAGGGCATGAACATGAAGGAGATCAAGGGCCGCACCGTGGCCGTGGACTG 563

Qy        141 ---------TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArg-ArgPr 157
                     ||||||  |||          :::::::::|||||||
Db        564 GGCCGTGGCCAAGGACAAGTACAAGGACACCCAGAGCGTGAGCGCCATCGGCGAGGAGAA 623

Qy        157 oSerLysLysLysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLy 177
              |||  :::   :::|||   :::        |||         ||||||:::
Db        624 GAGCCACGAGAGCAAGCACCAGGAGAGCGTGAAGAAGAAGGGCCGCGAGGAGGAGGACAT 683

Qy        177 sPheAspGluLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGl 197
               :::|||:::
Db        684 GGAGGAGGAGGAGAACGAC------------------------------------- 702

Qy        197 yGlnProValAlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGl 217
                          :::          :::   ||||||   |||:::||||
Db        703 -GACGACGACGACGACGACGACGAGGAGGACGGCGTGTTCGACGACGAGGACGAGGAGGA 761

Qy        217 uProAspLeuLysThrGlyLeu--------------TyrSerLysArgAlaAla----- 230
              |   ::::::::::::   :::                 ||||||||
Db        762 GGAGAACATCGAGAGCAAGGTGACCAAGCCCGTGCAGATCCAGAAGCGCGCCGTGAAGCG 821

Qy        231 ----------AlaLysSerAspAspThrSerAspAspAspPhe---MetGluGluGlyGl 246
                       |||||||||   |||   |||:::::|||      :::||||||
Db        822 CCCCGCCCCCGCCAAGAGCAGCGACCACAGCGAGGAGGACAGCGACCTGGAGGAGAGCGA 881

Qy        246 yGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySerGluPheLeuGlnArgAs 266
                |||                          ||||||   |||      |||   ||
Db        882 CAGCATCGAC----------------------GACGGCGAGGAGCTGGCCCAGAGCGA 917

Qy        266 pPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGlnAsnMetSerLysGln-- 285
              |   |||        :::                  :::   |||      ||:::
Db        918 CACCAGCACCGAGGAGCAGGAGGACAAGGCCGTGCAGGTGAGCAACAAGAAGAAGCGCAA 977

Qy        285 ------------------------------------------------------------ 285

Db        978 GCTGCCCAGCGACGTGAACGAGGGCAAGACCGTGTTCATCCGCAACCTGAGCTTCGACAG 1037

Qy        286 ------------------GluLeuIleLysGluTyrLeuGluLeu-------------- 294
                              ||||||:::::::::::   ||||||
Db       1038 CGAGGAGGAGGAGCTGGGCGAGCTGCTGCAGCAGTTCGGCGAGCTGAAGTACGTGCGCAT 1097

Qy        294 ------------------------------------------------------------ 294

Db       1098 CGTGCTGCACCCCGACACCGAGCACAGCAAGGGCTGCGCCTTCGCCCAGTTCATGACCCA 1157

Qy        295 ----------GluLysCysLeu------SerArgMetGluAspGluAsnAsnArgLeuAr 309
                        :::|||||||||      :::      |||:::|||         |||::
Db       1158 GGAGGCCGCCCAGAAGTGCCTGCTGGCCGCCAGCCCCGAGAACGAGGCCGGCGGCCTGAA 1217

Qy        309 gLeuGluSerLysArgLeuGlyGlyAspAspAlaArgValArgGluLeuGluLeuGluLe 329
               :|||:::   ::::::|||                          ::::::::::|||   ::
Db       1218 GCTGGACGGCCGCCAGCTG----------------------AAGGTGGACCTGGCCGT 1253

Qy        329 uAspArgLeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnGluLeuHisArgGlnGl 349
               :  |||      |||     ||||||             |||   |||||
Db       1254 GACCCGCGACGAGGCCGCCAAGCTGCAGACCACCAAGGT-GAAGAAGCCCACCGGCACCC 1312

Qy        349 nGluArgAlaPro 353
                 :::      |||
Db       1313 GCAACCTGTACCT 1325
```

RESULT 4
US-08-676-974-5
; Sequence 5, Application US/08676974
; Patent No. 5770422
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,974
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-676-974-5

Alignment Scores:

Pred. No.:	0.00055	Length:	2277
Score:	155.00	Matches:	96
Percent Similarity:	33.93%	Conservative:	55
Best Local Similarity:	21.57%	Mismatches:	147
Query Match:	8.12%	Indels:	148
DB:	1	Gaps:	18

US-09-972-758A-2 (1-359) x US-08-676-974-5 (1-2277)

```
Qy          6 LeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGlyAlaAlaAlaValGln 25
              |||   :::  ||||||                          |||      |||   :::
Db        144 CTACGTGACCTTCAGCAT--------------------GCTGGAGGACGTGCAGCGCGC 182
```

```
Qy         26 GluGlu--------------------LeuAsnProGluArgProProGlyAlaGluGlu 38
Db        183 CCTGAAGGAGATCACCACCTTCGAGGGCTGCAAGATCAACGTGACCGTGGCCAAGAAGAA 242

Qy         39 ArgValProGluGluAspSerArgTrpGlnSerArg------AlaPheProGlnLeuGly 56
Db        243 GCTGCGCAACAAGACCAAGGAGAAGGGCAAGAACGAGAACAGCGAGTGCCCCAAGAAGGA 302

Qy         57 GlyArgProGlyProGluGlyGluGlySerLeuGluSerGlnPro-------------- 71
Db        303 GCCCAA---GGCCAAGAAGGCCAAGGTGGCCGACAAGAAGGCCCGCCTGATCATCCGCAA 359

Qy         72 ProProLeuGlnThrGlnAlaCysProGluSerSerCysLeuArgGluGlyGluLysGly 91
Db        360 CCTGAGCTTCAAGTGCAGCGAGGACGACCTGAAGACCGTGTTCGCCCAGTTCGGCGC--- 416

Qy         92 GlnAsnGlyAspAspSerSerAlaGlyGlyAspPheProProProAlaGluValGluPro 111
Db        417 -----------------CGTGCTGGAGGTGAACATCCCCCGCAAGCCCGACGGCAA--- 455

Qy        112 ThrProGluAlaGluLeuLeuAlaGlnProCysHisAspSerGluAlaSerLysLeuGly 131
Db        456 ------GATGCGCGGCTTCGGCTTCGTGCAGTTCAAGAACCTGCTGGAGGCCGGCAAGGC 509

Qy        132 AlaProAlaAlaGlyGlyGluGluGlu------------------------------- 140
Db        510 ------CCTGAAGGGCATGAACATGAAGGAGATCAAGGGCCGCACCGTGGCCGTGGACTG 563

Qy        141 ---------TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArg-ArgPr 157
Db        564 GGCCGTGGCCAAGGACAAGTACAAGGACACCCAGAGCGTGAGCGCCATCGGCGAGGAGAA 623

Qy        157 oSerLysLysLysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLy 177
Db        624 GAGCCACGAGAGCAAGCACCAGGAGAGCGTGAAGAAGAAGGGCCGCGAGGAGGAGGACAT 683

Qy        177 sPheAspGluLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGl 197
Db        684 GGAGGAGGAGGAGAACGAC---------------------------------------- 702

Qy        197 yGlnProValAlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGl 217
Db        703 -GACGACGACGACGACGACGAGGAGGACGGCGTGTTCGACGACGAGGACGAGGAGGA 761

Qy        217 uProAspLeuLysThrGlyLeu--------------TyrSerLysArgAlaAla----- 230
Db        762 GGAGAACATCGAGAGCAAGGTGACCAAGCCCGTGCAGATCCAGAAGCGCGCCGTGAAGCG 821

Qy        231 ----------AlaLysSerAspAspThrSerAspAspAspPhe---MetGluGluGlyGl 246
Db        822 CCCCGCCCCCGCCAAGAGCAGCGACCACAGCGAGGAGGACAGCGACCTGGAGGAGAGCGA 881

Qy        246 yGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySerGluPheLeuGlnArgAs 266
Db        882 CAGCATCGAC----------------------GACGGCGAGGAGCTGGCCCAGAGCGA 917

Qy        266 pPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGlnAsnMetSerLysGln-- 285
Db        918 CACCAGCACCGAGGAGCAGGAGGACAAGGCCGTGCAGGTGAGCAACAAGAAGAAGCGCAA 977

Qy        285 ------------------------------------------------------------ 285
Db        978 GCTGCCCAGCGACGTGAACGAGGGCAAGACCGTGTTCATCCGCAACCTGAGCTTCGACAG 1037

Qy        286 -------------------GluLeuIleLysGluTyrLeuGluLeu-------------- 294
Db       1038 CGAGGAGGAGGAGCTGGGCGAGCTGCTGCAGCAGTTCGGCGAGCTGAAGTACGTGCGCAT 1097

Qy        294 ------------------------------------------------------------ 294
Db       1098 CGTGCTGCACCCCGACACCGAGCACAGCAAGGGCTGCGCCTTCGCCCAGTTCATGACCCA 1157
```

```
Qy        295 ----------GluLysCysLeu------SerArgMetGluAspGluAsnAsnArgLeuAr 309
Db       1158 GGAGGCCGCCCAGAAGTGCCTGCTGGCCGCCAGCCCCGAGAACGAGGCCGGCGGCCTGAA 1217

Qy        309 gLeuGluSerLysArgLeuGlyGlyAspAspAlaArgValArgGluLeuGluLeuGluLe 329
Db       1218 GCTGGACGGCCGCCAGCTG----------------------AAGGTGGACCTGGCCGT 1253

Qy        329 uAspArgLeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnGluLeuHisArgGlnGl 349
Db       1254 GACCCGCGACGAGGCCGCCAAGCTGCAGACCACCAAGGT-GAAGAAGCCCACCGGCACCC 1312

Qy        349 nGluArgAlaPro 353
Db       1313 GCAACCTGTACCT 1325
```

RESULT 5
US-09-098-487-5
; Sequence 5, Application US/09098487
; Patent No. 5917025
; GENERAL INFORMATION:
; APPLICANT: COLLINS, Kathleen
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,487
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-098-487-5

Alignment Scores:

Pred. No.:	0.00055	Length:	2277
Score:	155.00	Matches:	96
Percent Similarity:	33.93%	Conservative:	55
Best Local Similarity:	21.57%	Mismatches:	147
Query Match:	8.12%	Indels:	148
DB:	2	Gaps:	18

US-09-972-758A-2 (1-359) x US-09-098-487-5 (1-2277)

```
Qy          6 LeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGlyAlaAlaAlaValGln 25
Db        144 CTACGTGACCTTCAGCAT--------------------GCTGGAGGACGTGCAGCGCGC 182

Qy         26 GluGlu--------------------LeuAsnProGluArgProProGlyAlaGluGlu 38
Db        183 CCTGAAGGAGATCACCACCTTCGAGGGCTGCAAGATCAACGTGACCGTGGCCAAGAAGAA 242
```

```
Qy        39 ArgValProGluGluAspSerArgTrpGlnSerArg------AlaPheProGlnLeuGly 56
Db       243 GCTGCGCAACAAGACCAAGGAGAAGGGCAAGAACGAGAACAGCGAGTGCCCCAAGAAGGA 302

Qy        57 GlyArgProGlyProGluGlyGluGlySerLeuGluSerGlnPro--------------- 71
Db       303 GCCCAA---GGCCAAGAAGGCCAAGGTGGCCGACAAGAAGGCCCGCCTGATCATCCGCAA 359

Qy        72 ProProLeuGlnThrGlnAlaCysProGluSerSerCysLeuArgGluGlyGluLysGly 91
Db       360 CCTGAGCTTCAAGTGCAGCGAGGACGACCTGAAGACCGTGTTCGCCCAGTTCGGCGC--- 416

Qy        92 GlnAsnGlyAspAspSerSerAlaGlyGlyAspPheProProProAlaGluValGluPro 111
Db       417 -----------------CGTGCTGGAGGTGAACATCCCCCGCAAGCCCGACGGCAA--- 455

Qy       112 ThrProGluAlaGluLeuLeuAlaGlnProCysHisAspSerGluAlaSerLysLeuGly 131
Db       456 ------GATGCGCGGCTTCGGCTTCGTGCAGTTCAAGAACCTGCTGGAGGCCGGCAAGGC 509

Qy       132 AlaProAlaAlaGlyGlyGluGluGlu------------------------------- 140
Db       510 ------CCTGAAGGGCATGAACATGAAGGAGATCAAGGGCCGCACCGTGGCCGTGGACTG 563

Qy       141 ---------TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArg-ArgPr 157
Db       564 GGCCGTGGCCAAGGACAAGTACAAGGACACCCAGAGCGTGAGCGCCATCGGCGAGGAGAA 623

Qy       157 oSerLysLysLysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLy 177
Db       624 GAGCCACGAGAGCAAGCACCAGGAGAGCGTGAAGAAGAAGGGCCGCGAGGAGGAGGACAT 683

Qy       177 sPheAspGluLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGl 197
Db       684 GGAGGAGGAGGAGAACGAC---------------------------------------- 702

Qy       197 yGlnProValAlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGl 217
Db       703 -GACGACGACGACGACGACGACGAGGAGGACGGCGTGTTCGACGACGAGGACGAGGAGGA 761

Qy       217 uProAspLeuLysThrGlyLeu---------------TyrSerLysArgAlaAla----- 230
Db       762 GGAGAACATCGAGAGCAAGGTGACCAAGCCCGTGCAGATCCAGAAGCGCGCCGTGAAGCG 821

Qy       231 ----------AlaLysSerAspAspThrSerAspAspAspPhe---MetGluGluGlyGl 246
Db       822 CCCCGCCCCCGCCAAGAGCAGCGACCACAGCGAGGAGGACAGCGACCTGGAGGAGAGCGA 881

Qy       246 yGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySerGluPheLeuGlnArgAs 266
Db       882 CAGCATCGAC-----------------------GACGGCGAGGAGCTGGCCCAGAGCGA 917

Qy       266 pPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGlnAsnMetSerLysGln-- 285
Db       918 CACCAGCACCGAGGAGCAGGAGGACAAGGCCGTGCAGGTGAGCAACAAGAAGAAGCGCAA 977

Qy       285 ------------------------------------------------------------ 285
Db       978 GCTGCCCAGCGACGTGAACGAGGGCAAGACCGTGTTCATCCGCAACCTGAGCTTCGACAG 1037

Qy       286 -------------------GluLeuIleLysGluTyrLeuGluLeu-------------- 294
Db      1038 CGAGGAGGAGGAGCTGGGCGAGCTGCTGCAGCAGTTCGGCGAGCTGAAGTACGTGCGCAT 1097

Qy       294 ------------------------------------------------------------ 294
Db      1098 CGTGCTGCACCCCGACACCGAGCACAGCAAGGGCTGCGCCTTCGCCCAGTTCATGACCCA 1157

Qy       295 ----------GluLysCysLeu------SerArgMetGluAspGluAsnAsnArgLeuAr 309
Db      1158 GGAGGCCGCCCAGAAGTGCCTGCTGGCCGCCAGCCCCGAGAACGAGGCCGGCGGCCTGAA 1217

Qy       309 gLeuGluSerLysArgLeuGlyGlyAspAspAlaArgValArgGluLeuGluLeuGluLe 329
Db      1218 GCTGGACGGCCGCCAGCTG----------------------AAGGTGGACCTGGCCGT 1253

Qy       329 uAspArgLeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnGluLeuHisArgGlnGl 349
Db      1254 GACCCGCGACGAGGCCGCCAAGCTGCAGACCACCAAGGT-GAAGAAGCCCACCGGCACCC 1312

Qy       349 nGluArgAlaPro 353
Db      1313 GCAACCTGTACCT 1325
```

RESULT 6
US-09-620-312D-248
; Sequence 248, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 248
; LENGTH: 7453
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(4362)
US-09-620-312D-248

Alignment Scores:

Pred. No.:	0.0136	Length:	7453
Score:	147.00	Matches:	103
Percent Similarity:	36.84%	Conservative:	51
Best Local Similarity:	24.64%	Mismatches:	170
Query Match:	7.70%	Indels:	95
DB:	4	Gaps:	15

US-09-972-758A-2 (1-359) x US-09-620-312D-248 (1-7453)

```
Qy        15 ThrSerAsnCysThrGlyAlaAlaAlaValGlnGlu------GluLeuAsnProGluArg 32
Db      3948 ACTACAGGATACACAGGAGCTTCTTCAGGAGGAGACACGCCAGAAACTAAACCTGAGCAG 4007

Qy        33 ProProGlyAlaGluGluArgValProGluGluAspSerArgTrpGlnSerArgAlaPhe 52
Db      4008 TCGGATCCGGCAGCTGGAAGAGGAGAAGAACAGTCTTCA-------------------- 4046

Qy        53 ProGlnLeuGlyGlyArgProGlyProGluGlyGluGlySerLeuGluSerGlnProPro 72
```

```
                  ||||||      |||   :::   |||            |||      |||
Db       4047 GGAGCAGCAGGAGGAGGAGGAGGAGGCCAGGAAGAACCTGGAGAAGCAAGTGCTGGCCCT 4106

Qy         73 ProLeuGlnThrGlnAlaCysProGluSerSerCysLeuArgGluGly---------Glu  89
              :::                  |||   |||      |||   |||          :::
Db       4107 GCAGTCCCAGTTGGCTGATACCAAGAAGAAAGTAGATGACGACCTGGGAACAATTGAAAG 4166

Qy         90 LysGlyGlnAsnGlyAspAspSerSerAlaGlyGlyAspPheProProProAlaGluVal 109
              |||::::::   :::::::::|||   |||          |||   ||||||
Db       4167 TCTGGAAGAAGCCAAGAAGAAGCTTCTGAAGGACGCGGAGGCCCTGAGCCAGCG------ 4220

Qy        110 GluProThrProGluAlaGluLeuLeuAlaGlnProCysHisAspSerGluAlaSerLys 129
              |||      |||       :::   |||      |||   |||
Db       4221 ---CCTGGAGGAGAAGGCACTGGCGTATGACAAACTGGAGAAGACCAAGAACCGCCTGCA 4277

Qy        130 LeuGlyAla----------------------ProAlaAlaGlyGlyGluGluGluTrp 141
              ||||||                       |||   |||   |||      :::
Db       4278 GCAGGAGCTGGACGACCTCACGGTGGACCTGGACCACCAGCGCCAGGTCGCCTCCAACTT 4337

Qy        142 GlyGlnGlnGlnArgGln---------------------------------------- 147
              |||::::::   |||:::
Db       4338 GGAGAAGAAGCAAGAAGAAGTTTGACCAGCTGTTAGCAGAAGAGAAGAGCATCTCTGCTC 4397

Qy        148 ------------------LeuGlyLysLysLysHisArgArgArgProSerLysLysLys 161
                               |||   |||          |||      |||
Db       4398 GCTATGCCGAAGAGCGGGACCGGGCCGAAGCCGAGGCCAGAGAGAAAGAAACCAAAGCCC 4457

Qy        162 ArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAsp---Glu 180
              ||||||           :::      |||      |||::::::|||:::   :::
Db       4458 TGTCACTGGCCCGGGCCCTCGAGGAAGCCCTGGAGGCC-AAGGAGGAGTTTGAGAGGCAG 4516

Qy        181 LysGlnSerLeuArgAlaSer---------------------------------------- 187
              :::   |||||||||
Db       4517 AACAAGCAGCTCCGAGCAGACATGGAAGACCTCATGAGCTCCAAAGATGATGTGGGAAAA 4576

Qy        188 ---------ArgIleArgAlaGluMetPheAlaLysGlyGlnProValAlaProTyrAsn 204
                      :::          |||      |||   |||
Db       4577 AACGTTCACGAACTTGAAAAATCCAAACGGGCCCTAGAGCAGCAGGTGGAGGAAATGAGG 4636

Qy        205 ThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeuLysThrGlyLeu 224
              |||          ::::::|||      |||          |||   |||      |||
Db       4637 ACCCAGCTGGAGGAGCTGGAAGACGAACTCCAGGCCACGGAAGATGCCAAGCTTCGTCTG 4696

Qy        225 TyrSerLysArgAlaAlaAlaLysSerAsp---------------AspThrSerAspAsp 239
                       |||   |||:::                      |||   |||:::
Db       4697 GAGGTCAACATGCAGGCCATGAAGGCGCAGTTCGAGAGAGACCTGCAAACCAGGGATGAG 4756

Qy        240 --AspPheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspG 259
                :::      ||||||          :::      |||:::          ||||||   |
Db       4757 CAGAATGAAGAGAAGAAGCGGCTGCTGATCAAACAGGTGCGGGAGCTCGAGGCGGAGCTG 4816

Qy        259 ly-SerGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSer 278
              ||    |||      ||||||      :::          ::::::
Db       4817 GAGGATGAGAGGAAACAGCGGGCGCTTGCTGTAGCTTCAAAGAAAAAGATGGAGATAGAC 4876

Qy        279 LeuGln--------------------------AsnMetSerLysGlnGluLeuIleLys 289
              |||:::                        |||   ::::::   |||:::||||||
Db       4877 CTGAAGGACCTCGAAGCCCAAATCGAGGCTGCGAACAAAGCTCGGGATGAGGTGATTAAG 4936

Qy        290 GluTyrLeuGluLeuGluLysCysLeuSerArgMetGluAspGluAsnAsnArgLeuArg 309
                   :::|||   |||      ::::::|||:::|||
Db       4937 ---------CAGCTCCGCAAGCTCCAGGCTCAGATGAAGGATTACCAACGT------GAA 4981

Qy        310 LeuGluSerLysArgLeuGlyGlyAspAsp------AlaArgValArgGluLeuGluLeu 327
              ||||||      |||      |||:::      |||:::   :::|||   |||
Db       4982 TTAGAAGAAGCTCGTGCATCCAGAGATGAGATTTTTGCTCAATCCAAAGAGAGTGAAAAG 5041

Qy        328 GluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnGluLeu------ 345
              :::|||      |||   ||||||   |||||||||      :::   ||||||
Db       5042 AAATTGAAGAGTCTGGAAGCAGAAATCCTTCAATTG------CAGGAGGAACTTGCCTCA 5095

Qy        346 -----------------HisArgGlnGlnGluArgAlaProLeuSer 355
                              |||   :::|||||||||      |||:::
```

```
Db       5096 TCTGAGCGAGCCCGCCGACACGCCGAGCAGGAGAGAGATGAGCTGGCG 5143

RESULT 7
US-09-620-312D-249
; Sequence 249, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Chen, Rui-hong
;  APPLICANT:  Zhao, Qing A.
;  APPLICANT:  Wehrman, Tom
;  APPLICANT:  Xue, Aidong J.
;  APPLICANT:  Yang, Yonghong
;  APPLICANT:  Wang, Jian-Rui
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Ma, Yunqing
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Wang, Zhiwei
;  APPLICANT:  John Tillinghast
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. 6569662el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 784CIP2B
;  CURRENT APPLICATION NUMBER: US/09/620,312D
;  CURRENT FILING DATE:  2000-07-19
;  PRIOR APPLICATION NUMBER: 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 1105
;  SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 249
;   LENGTH: 7501
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(4410)
US-09-620-312D-249

Alignment Scores:
Pred. No.:                0.0137          Length:          7501
Score:                    147.00          Matches:         103
Percent Similarity:       36.84%          Conservative:    51
Best Local Similarity:    24.64%          Mismatches:      170
Query Match:              7.70%           Indels:          95
DB:                       4               Gaps:            15

US-09-972-758A-2 (1-359) x US-09-620-312D-249 (1-7501)

Qy         15 ThrSerAsnCysThrGlyAlaAlaAlaValGlnGlu------GluLeuAsnProGluArg 32
              |||:::      |||||||||:::::      :::      |||      ||||||:::
Db       3996 ACTACAGGATACACAGGAGCTTCTTCAGGAGGAGACACGCCAGAAACTAAACCTGAGCAG 4055

Qy         33 ProProGlyAlaGluGluArgValProGluGluAspSerArgTrpGlnSerArgAlaPhe 52
                        |||      |||      |||:::   |||
Db       4056 TCGGATCCGGCAGCTGGAAGAGGAGAAGAACAGTCTTCA---------------------- 4094

Qy         53 ProGlnLeuGlyGlyArgProGlyProGluGlyGluGlySerLeuGluSerGlnProPro 72
                       ||||||      |||   :::   |||          |||      |||
Db       4095 GGAGCAGCAGGAGGAGGAGGAGGAGGCCAGGAAGAACCTGGAGAAGCAAGTGCTGGCCCT 4154

Qy         73 ProLeuGlnThrGlnAlaCysProGluSerSerCysLeuArgGluGly---------Glu  89
              :::                  |||   |||      |||   |||          :::
Db       4155 GCAGTCCCAGTTGGCTGATACCAAGAAGAAAGTAGATGACGACCTGGGAACAATTGAAAG 4214

Qy         90 LysGlyGlnAsnGlyAspAspSerSerAlaGlyGlyAspPheProProProAlaGluVal 109
              |||::::::   :::::::::|||   |||          |||   ||||||
Db       4215 TCTGGAAGAAGCCAAGAAGAAGCTTCTGAAGGACGCGGAGGCCCTGAGCCAGCG------ 4268
```

```
Qy        110 GluProThrProGluAlaGluLeuLeuAlaGlnProCysHisAspSerGluAlaSerLys 129
Db       4269 ---CCTGGAGGAGAAGGCACTGGCGTATGACAAACTGGAGAAGACCAAGAACCGCCTGCA 4325

Qy        130 LeuGlyAla-----------------------ProAlaAlaGlyGlyGluGluGluTrp 141
Db       4326 GCAGGAGCTGGACGACCTCACGGTGGACCTGGACCACCAGCGCCAGGTCGCCTCCAACTT 4385

Qy        142 GlyGlnGlnGlnArgGln---------------------------------------- 147
Db       4386 GGAGAAGAAGCAAGAAGAAGTTTGACCAGCTGTTAGCAGAAGAGAAGAGCATCTCTGCTC 4445

Qy        148 ------------------LeuGlyLysLysLysHisArgArgArgProSerLysLysLys 161
Db       4446 GCTATGCCGAAGAGCGGGACCGGGCCGAAGCCGAGGCCAGAGAGAAAGAAACCAAAGCCC 4505

Qy        162 ArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAsp---Glu 180
Db       4506 TGTCACTGGCCCGGGCCCTCGAGGAAGCCCTGGAGGCC-AAGGAGGAGTTTGAGAGGCAG 4564

Qy        181 LysGlnSerLeuArgAlaSer------------------------------------- 187
Db       4565 AACAAGCAGCTCCGAGCAGACATGGAAGACCTCATGAGCTCCAAAGATGATGTGGGAAAA 4624

Qy        188 ---------ArgIleArgAlaGluMetPheAlaLysGlyGlnProValAlaProTyrAsn 204
Db       4625 AACGTTCACGAACTTGAAAAATCCAAACGGGCCCTAGAGCAGCAGGTGGAGGAAATGAGG 4684

Qy        205 ThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeuLysThrGlyLeu 224
Db       4685 ACCCAGCTGGAGGAGCTGGAAGACGAACTCCAGGCCACGGAAGATGCCAAGCTTCGTCTG 4744

Qy        225 TyrSerLysArgAlaAlaAlaLysSerAsp---------------AspThrSerAspAsp 239
Db       4745 GAGGTCAACATGCAGGCCATGAAGGCGCAGTTCGAGAGAGACCTGCAAACCAGGGATGAG 4804

Qy        240 --AspPheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspG 259
Db       4805 CAGAATGAAGAGAAGAAGCGGCTGCTGATCAAACAGGTGCGGGAGCTCGAGGCGGAGCTG 4864

Qy        259 ly-SerGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSer 278
Db       4865 GAGGATGAGAGGAAACAGCGGGCGCTTGCTGTAGCTTCAAAGAAAAAGATGGAGATAGAC 4924

Qy        279 LeuGln--------------------------AsnMetSerLysGlnGluLeuIleLys 289
Db       4925 CTGAAGGACCTCGAAGCCCAAATCGAGGCTGCGAACAAAGCTCGGGATGAGGTGATTAAG 4984

Qy        290 GluTyrLeuGluLeuGluLysCysLeuSerArgMetGluAspGluAsnAsnArgLeuArg 309
Db       4985 ---------CAGCTCCGCAAGCTCCAGGCTCAGATGAAGGATTACCAACGT------GAA 5029

Qy        310 LeuGluSerLysArgLeuGlyGlyAspAsp------AlaArgValArgGluLeuGluLeu 327
Db       5030 TTAGAAGAAGCTCGTGCATCCAGAGATGAGATTTTTGCTCAATCCAAAGAGAGTGAAAAG 5089

Qy        328 GluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnGluLeu------ 345
Db       5090 AAATTGAAGAGTCTGGAAGCAGAAATCCTTCAATTG------CAGGAGGAACTTGCCTCA 5143

Qy        346 ------------------HisArgGlnGlnGluArgAlaProLeuSer 355
Db       5144 TCTGAGCGAGCCCGCCGACACGCCGAGCAGGAGAGAGATGAGCTGGCG 5191
```

RESULT 8
US-09-252-991A-2538
; Sequence 2538, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2538
; LENGTH: 2820
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2538

Alignment Scores:

Pred. No.:	0.00946	Length:	2820
Score:	142.00	Matches:	98
Percent Similarity:	31.26%	Conservative:	48
Best Local Similarity:	20.99%	Mismatches:	156
Query Match:	7.43%	Indels:	165
DB:	4	Gaps:	19

US-09-972-758A-2 (1-359) x US-09-252-991A-2538 (1-2820)

```
Qy         12 GlnProGlnThrSerAsnCysThrGlyAlaAlaAlaValGlnGluGluLeuAsnProGlu 31
Db        139 CAACCTGAAGCGGATCGGCGAACTGCT--------------CGCCGACTACGACCTCGT 183

Qy         32 ArgProProGlyAlaGluGluArgValProGluGluAspSerArgTrpGlnSerArgAla 51
Db        184 CGCCCTCCAGGAAGCCGACGGCGGCAGCCTGCGTTCCGGCTACGTCAATCAGGTCGAGCA 243

Qy         52 PheProGlnLeuGlyGly-----------------------ArgProGlyProGluGly 63
Db        244 CTTGGCCCAGTTGGGGGCGTTCCCCTACTGGTACCAGCAGCTCAACCGCAACCTCGGGCG 303

Qy         64 GluGlySerLeuGluSerGlnPro---ProProLeuGlnThrGlnAlaCysProGluSer 82
Db        304 TATCGCCCAGCACAGCAACGGCCTGCTCAGCCGCTGGCGGCCGACCCTGCTGGAAGATCA 363

Qy         83 SerCysLeuArgGluGly---------------------------------------- 88
Db        364 TCCGTTGCCCGGTCCGCCGGGACGCGGCGCGATCTTCGTGCGCTTTGGCGAAGGCGCCGA 423

Qy         89 GluLysGlyGlnAsnGlyAspAspSerSerAlaGlyGlyAspPhePro---ProProAla 107
Db        424 CGCGCTGGTGGTAGTGATGATGCATCTCGCTCTCGGCGCTCGCACCCGCACCCGCCAGCT 483

Qy        108 GluValGluProThrProGluAlaGluLeu---------------------------- 117
Db        484 CGCCTATATCCGCGAGCTGATCGGCGACTACCGCCATCAGGTGCTGATGGGCGACATGAA 543

Qy        118 --------------------------------LeuAlaGlnProCysHisAspSerGlu 126
Db        544 CACCCACGCCGTCGATCTTCTGGAAAACTCGCCCCTGCGCGACCTCGGCCTGATGGCGCC 603

Qy        127 Ala---------------SerLysLeuGlyAlaProAla--------------------- 134
Db        604 GCAGGTCGAGGCGACGTTTCCCAGCTGGCGGCCGCAGCGCAATCTCGATCACATCCTGCT 663

Qy        134 ------------------------------------------------------------ 134
Db        664 CAGTTCCGATCTCGCTCTCGAGCGGGTCGACGTGCTCGACCAGCCGATCTCCGACCATCT 723

Qy        135 ------AlaGlyGlyGluGluGluTrpGlyGlnGlnGlnArgGln---LeuGlyLysLys 151
Db        724 GCCGGTATCGGTGGAGATTCGCCTGCCGGACAACCTGCGCGACAAGGCCTTGGTCGTGCG 783

Qy        152 LysHisArgArgArgProSerLysLysLys---------ArgHisTrpLysProTyrTyr 168
Db        784 CGGGGATGCCGGGAGTCCGCCATGAGTCGCGACGACGTCCAGCGCTGGAAAGACAAGTAC 843

Qy        169 LysLeuThrTrpGluGluLysLysLysPheAspGluLysGlnSerLeuArgAlaSerArg 188
```

```
Db        844 CTGGAGAACATCGAGCAGCAGGAGCGCCTGCAACGGCGCTGGGATGCCCGGATCGACCTG 903

Qy        189 IleArgAlaGluMetPheAlaLysGlyGlnProValAlaProTyrAsnThrThrGlnPhe 208
Db        904 CTGCGT---------------------------------------------------------- 909

Qy        209 LeuMetAspAspHisAspGlnGluGluProAspLeuLysThrGlyLeuTyrSerLysArg 228
Db        910 ----------------------------------------CGCGGCCTGGTGCGTAGCAGT 930

Qy        229 AlaAlaAlaLysSerAspAspThrSerAspAspAspPheMetGluGlu------------ 244
Db        931 CTGGCGGCCGAAGGCAGCGACAAGGCGGTCGACCAGTGCATGAAGGAGCTGCGCGAGATC 990

Qy        245 ------GlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySerGluPhe 262
Db        991 CTCCGTCGCGACGACATGGACGCCGGCCTGTCGGGCCTG------------ATTCCGCGC 1038

Qy        263 LeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGlnAsnMet 282
Db       1039 CTGGAAAAGGCCGTGCTCGACTCCGAACAGCGCCGCCAGCAACGCACCCAGCAGAACATC 1098

Qy        283 ---SerLysGlnGluLeuIleLysGluTyrLeu------------------GluLeuGlu 295
Db       1099 GATGCTCTCGGCGAACTGGCCCAGCAGTTGCTGGCCCTGGACCTGCCGCGGGAACTGCGC 1158

Qy        296 LysCysLeuSerArg---------------------------------------------- 300
Db       1159 AAGCCGCTGAAGCAGTTCGCCCGCGATATCGAGGAGCGCGCCCGGCAGTCGCGCGAGATT 1218

Qy        301 ------MetGluAspGluAsnAsnArgLeuArgLeuGlu------SerLysArgLeuGly 316
Db       1219 CCGATCCTGCTCAGCGAGCTGAGCAGGCTGCAACGCCAGGCGCTGGCCGAGCGCAAGGGC 1278

Qy        317 GlyAsp------AspAlaArgValArgGluLeu--------------------GluLeu 327
Db       1279 GGCGATGCCGAAGACGGTCGCCCGAGCCTGTTGCAACGCCTGTTCGGCGGCAAGGAGTCG 1338

Qy        328 GluLeu-AspArgLeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnGluLeuHisAr 347
Db       1339 GAAACCACCGCGGAGCCGAGCGCATCGGTGCCGTCCGTTGTGGCGGCAAGCAATACACCG 1398

Qy        347 gGlnGlnGluArgAlaPro 353
Db       1399 ATCCAACCGGCGGCGGCCG 1417

RESULT 9
US-08-464-266-1
; Sequence 1, Application US/08464266
; Patent No. 5641652
;  GENERAL INFORMATION:
;    APPLICANT:  ORO, Ph.D., ANTHONY E.
;    APPLICANT:  EVANS, Ph.D., RONALD M.
;    TITLE OF INVENTION:  INSECT RETINOID-LIKE RECEPTOR
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  444 South Flower Street, Suite 2000
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  United States
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/464,266
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/013,975
;      FILING DATE:  04-FEB-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/497,935
;      FILING DATE:  22-FEB-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter,, Stephen E.
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:  P41 9966
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-546-4737
;      TELEFAX:  619-546-9392
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2304 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  163..1704
US-08-464-266-1

Alignment Scores:
Pred. No.:              0.00871      Length:          2304
Score:                  141.00       Matches:         106
Percent Similarity:     32.43%       Conservative:    49
Best Local Similarity:  22.18%       Mismatches:      127
Query Match:            7.38%        Indels:          197
DB:                     1            Gaps:            22

US-09-972-758A-2 (1-359) x US-08-464-266-1 (1-2304)

Qy         13 ProGlnThrSerAsnCysThrGlyAlaAlaAlaValGlnGluGluLeuAsnProGluArg 32
Db        388 CCCAATTCGGCTGGAGGCTCTGCCGCCGCTGCAGTCCAGCAGCAGTATCCGCCTAACCAT 447

Qy         33 ProProGlyAlaGluGluArg---------ValProGluGluAspSerArgTrpGlnSer 49
Db        448 CCGCTGAGCGGCAGCAAGCACCTCTGCTCTATTTG-CGGGGATCGGGCCAGTGGCAAGCA 506

Qy         50 ArgAlaPheProGlnLeuGlyGlyArgProGly------------ProGluGlyGluGly 65
Db        507 CTACGGCGTGTACAGCTGTGAGGGCTGCAAGGGCTTCTTTAAACGCACAGTGCGCAAGGA 566

Qy         66 Ser---------LeuGluSerGlnProProProLeuGlnThrGlnAlaCysProGluSer 82
Db        567 TCTCACATACGCTTGCAGGGAGAACCGCAACTGCATCATAGACAAGCGGCAGAGGAA--- 623

Qy         83 SerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGlyGlyAsp 102
Db        623 ---------------------------------------------------------- 623

Qy        103 PheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGlnProCys 122
Db        624 ---------CCGCTGCCAGTACTGCCGCTACCAGAAGTGCCTAACCTGCGG--------- 665

Qy        123 HisAspSerGluAlaSerLysLeuGlyAla------------ProAlaAlaGlyGlyGlu 138
Db        666 CATGAAGCGCGAAGCGGTCCAGGAGGAGCGTCAACGCGGCGCCCGCAATGCGGCGGG--- 722

Qy        139 GluGluTrpGlyGlnGlnGlnArg---------------------------------- 146
Db        723 ------TAGGCTCAGCGCCAGCGGAGGCGGCAGTAGCGGTCCAGGTTCGGTAGGCGGATC 776

Qy        147 GlnLeuGlyLysLysLysHisArgArgArgProSerLysLysLysArgHisTrpLysPro 166
Db        777 CAGCTCTCAAGGCGGAGGAGGAGGAGGCGG--------------CGTTTCTGGCGGAAT 821

Qy        167 TyrTyrLysLeuThrTrpGluGluLysLysLysPhe-AspGluLysGlnSerLeuArgAl 186
```

```
Db        822 ------------------GGGCAGCGGCAACGGTTCTGATGACTTCATGACCAATAGCGT 863

Qy        186 aSerArg----------------------IleArgAlaGluMetPheAlaLysGlyGln-- 198

Db        864 GTCCAGGGATTTCTCGATCGAGCGCATCATAGAGGCCGAGCAGCGAGCGGAGACCCAATG 923

Qy        199 -------------------------ProValAlaProTyrAsnThrThrGlnPheLeuMe 210

Db        924 CGGCGATCGTGCACTGACGTTCCTGCGCGTTGGTCCCTATTCCACAGTCCAG-------- 975

Qy        210 tAspAspHisAspGlnGluGluProAspLeuLysThrGlyLeu----------------- 224

Db        976 ----------------------CCGGACTACAAGGGTGCCGTGTCGGCCCTGTGCCAAGT 1013

Qy        225 -------------------------------TyrSerLysArgAlaAlaAlaLysSerAs 234

Db       1014 GGTCAACAAACAGCTCTTCCAGATGGTCGAATACGCGCGCATGATGCCGCACTTTGCCCA 1073

Qy        234 pAspThrSerAspAspAspPhe---------------------------------------- 241

Db       1074 GGTGCCGCTGGACGACCAGGTGATTCTGCTGAAAGCCGCTTGGATCGAGCTGCTCATTGC 1133

Qy        242 ----------------------------MetGluGluGlyGlyGluGluAspGlyGlySe 252

Db       1134 GAACGTGGCCTGGTGCAGCATCGTTTCGCTGGATGACGGCGGT---GCCGGCGGCGGGGG 1190

Qy        252 rAspGlyMetGlyGlyAspGlySer---------------------------------- 260

Db       1191 CGGTGGACTAGGCCACGATGGCTCCTTTGAGCGACGATCACCGGGCCTTCAGCCCCAGCA 1250

Qy        261 -GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeu-- 279

Db       1251 GCTGTTCCTCAACCAGAGCTTCTCG--------------TACCATCGCAACAGTGCGAT 1295

Qy        279 ------------------------------------------------------------ 279

Db       1296 CAAAGCCGGTGTGTCAGCCATCTTCGACCGCATATTGTCGGAGCTGAGTGTAAAGATGAA 1355

Qy        280 ----GlnAsnMetSerLysGlnGluLeu------------------------------- 287

Db       1356 GCGGCTGAATCTCGACCGACGCGAGCTGTCCTGCTTGAAGGCCATCATACTGTACAACCC 1415

Qy        288 ------------IleLysGluTyrLeuGluLeuGluLysCysLeuSerArgMetGlu-- 302

Db       1416 GGACATACGCGGGATCAAGAGCCGGGCGGAGATCGAGATGTGCCGCGAGAAGGTGTACGC 1475

Qy        303 -------AspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAl 320

Db       1476 TTGCCTGGACGAGCACTGCCGCCTGGAACATCCG--------------GGCGACGATGG 1520

Qy        320 aArgValArgGluLeuGluLeuGluLeuAspArg-LeuArgAlaGluAsnLeuGlnLeuL 340

Db       1521 ACGCTTTGCGCAACTGCTGCTGCGTCTGCGCCGCTTTGCGATCGATCAGCCTGAAGTGCC 1580

Qy        340 euThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLys 356

Db       1581 AGGATCACCTGTTCCTCTTCCGCATTACCAGCGACCGGCCGCTGGAGGAG 1630

RESULT 10
US-08-464-272-1
; Sequence 1, Application US/08464272
; Patent No. 5688691
;  GENERAL INFORMATION:
;    APPLICANT:  ORO, Ph.D., ANTHONY E.
;    APPLICANT:  EVANS, Ph.D., RONALD M.
;    TITLE OF INVENTION:  INSECT RETINOID-LIKE RECEPTOR
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  444 South Flower Street, Suite 2000
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  United States
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/464,272
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/013,975
;      FILING DATE:  04-FEB-1993
;      APPLICATION NUMBER:  US 07/497,935
;      FILING DATE:  22-FEB-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter,, Stephen E.
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:  P41 9350
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-546-4737
;      TELEFAX:  619-546-9392
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2304 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  163..1704
US-08-464-272-1

Alignment Scores:
Pred. No.:                0.00871        Length:          2304
Score:                    141.00         Matches:         106
Percent Similarity:       32.43%         Conservative:    49
Best Local Similarity:    22.18%         Mismatches:      127
Query Match:              7.38%          Indels:          197
DB:                       1              Gaps:            22

US-09-972-758A-2 (1-359) x US-08-464-272-1 (1-2304)

Qy         13 ProGlnThrSerAsnCysThrGlyAlaAlaAlaValGlnGluGluLeuAsnProGluArg 32

Db        388 CCCAATTCGGCTGGAGGCTCTGCCGCCGCTGCAGTCCAGCAGCAGTATCCGCCTAACCAT 447

Qy         33 ProProGlyAlaGluGluArg---------ValProGluGluAspSerArgTrpGlnSer 49

Db        448 CCGCTGAGCGGCAGCAAGCACCTCTGCTCTATTTG-CGGGGATCGGGCCAGTGGCAAGCA 506

Qy         50 ArgAlaPheProGlnLeuGlyGlyArgProGly------------ProGluGlyGluGly 65

Db        507 CTACGGCGTGTACAGCTGTGAGGGCTGCAAGGGCTTCTTTAAACGCACAGTGCGCAAGGA 566

Qy         66 Ser---------LeuGluSerGlnProProProLeuGlnThrGlnAlaCysProGluSer 82

Db        567 TCTCACATACGCTTGCAGGGAGAACCGCAACTGCATCATAGACAAGCGGCAGAGGAA--- 623

Qy         83 SerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGlyGlyAsp 102

Db        623 ------------------------------------------------------------ 623

Qy        103 PheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGlnProCys 122

Db        624 ---------CCGCTGCCAGTACTGCCGCTACCAGAAGTGCCTAACCTGCGG--------- 665

Qy        123 HisAspSerGluAlaSerLysLeuGlyAla------------ProAlaAlaGlyGlyGlu 138

Db        666 CATGAAGCGCGAAGCGGTCCAGGAGGAGCGTCAACGCGGCGCCCGCAATGCGGCGGG--- 722
```

```
Qy       139 GluGluTrpGlyGlnGlnGlnArg---------------------------------- 146
                 |||:::||||||
Db       723 ------TAGGCTCAGCGCCAGCGGAGGCGGCAGTAGCGGTCCAGGTTCGGTAGGCGGATC 776

Qy       147 GlnLeuGlyLysLysLysHisArgArgArgProSerLysLysLysArgHisTrpLysPro 166
             ||||||   :::::::::   |||||||||                |||   |||:::
Db       777 CAGCTCTCAAGGCGGAGGAGGAGGAGGCGG--------------CGTTTCTGGCGGAAT 821

Qy       167 TyrTyrLysLeuThrTrpGluGluLysLysLysPhe-AspGluLysGlnSerLeuArgAl 186
                                 :::::::::::||| |||:::     :::
Db       822 -----------------GGGCAGCGGCAACGGTTCTGATGACTTCATGACCAATAGCGT 863

Qy       186 aSerArg--------------------IleArgAlaGluMetPheAlaLysGlyGln-- 198
             ||||||                  |||   ||||||      |||:::   |||
Db       864 GTCCAGGGATTTCTCGATCGAGCGCATCATAGAGGCCGAGCAGCGAGCGGAGACCCAATG 923

Qy       199 ------------------------ProValAlaProTyrAsnThrThrGlnPheLeuMe 210
                                      |||   ||||||:::|||   |||
Db       924 CGGCGATCGTGCACTGACGTTCCTGCGCGTTGGTCCCTATTCCACAGTCCAG-------- 975

Qy       210 tAspAspHisAspGlnGluGluProAspLeuLysThrGlyLeu----------------- 224
                                    |||||   |||     :::
Db       976 ---------------------CCGGACTACAAGGGTGCCGTGTCGGCCCTGTGCCAAGT 1013

Qy       225 -------------------------------TyrSerLysArgAlaAlaAlaLysSerAs 234
                                           |||::::::                 :::
Db      1014 GGTCAACAAACAGCTCTTCCAGATGGTCGAATACGCGCGCATGATGCCGCACTTTGCCCA 1073

Qy       234 pAspThrSerAspAspAspPhe---------------------------------------- 241
                 ||||||
Db      1074 GGTGCCGCTGGACGACCAGGTGATTCTGCTGAAAGCCGCTTGGATCGAGCTGCTCATTGC 1133

Qy       242 ----------------------------MetGluGluGlyGlyGluGluAspGlyGlySe 252
                                        :::::::::||||||          ||||||
Db      1134 GAACGTGGCCTGGTGCAGCATCGTTTCGCTGGATGACGGCGGT---GCCGGCGGCGGGGG 1190

Qy       252 rAspGlyMetGlyGlyAspGlySer---------------------------------- 260
                ||||:::|||   |||||||||
Db      1191 CGGTGGACTAGGCCACGATGGCTCCTTTGAGCGACGATCACCGGGCCTTCAGCCCCAGCA 1250

Qy       261 -GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeu-- 279
                ||||||   :::   ||||||                ||||||      |||
Db      1251 GCTGTTCCTCAACCAGAGCTTCTCG--------------TACCATCGCAACAGTGCGAT 1295

Qy       279 ------------------------------------------------------------ 279

Db      1296 CAAAGCCGGTGTGTCAGCCATCTTCGACCGCATATTGTCGGAGCTGAGTGTAAAGATGAA 1355

Qy       280 ----GlnAsnMetSerLysGlnGluLeu------------------------------- 287
                    |||:::   :::::||||||
Db      1356 GCGGCTGAATCTCGACCGACGCGAGCTGTCCTGCTTGAAGGCCATCATACTGTACAACCC 1415

Qy       288 -------------IleLysGluTyrLeuGluLeuGluLysCysLeuSerArgMetGlu-- 302
                         ||||||          |||:::|||   |||      ::::::
Db      1416 GGACATACGCGGGATCAAGAGCCGGGCGGAGATCGAGATGTGCCGCGAGAAGGTGTACGC 1475

Qy       303 -------AspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAl 320
                     ||||||:::   ||||||                          |||||||||
Db      1476 TTGCCTGGACGAGCACTGCCGCCTGGAACATCCG--------------GGCGACGATGG 1520

Qy       320 aArgValArgGluLeuGluLeuGluLeuAspArg-LeuArgAlaGluAsnLeuGlnLeuL 340
              |||      :::|||   |||   |||   ||| ||||||:::   :::|||:::
Db      1521 ACGCTTTGCGCAACTGCTGCTGCGTCTGCGCCGCTTTGCGATCGATCAGCCTGAAGTGCC 1580

Qy       340 euThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLys 356
                         |||   |||                |||||   :::
Db      1581 AGGATCACCTGTTCCTCTTCCGCATTACCAGCGACCGGCCGCTGGAGGAG 1630

RESULT 11
US-08-464-514-1
; Sequence 1, Application US/08464514
; Patent No. 6265173
;  GENERAL INFORMATION:
;    APPLICANT:  EVANS, RONALD M.
;    APPLICANT:  MCKEOWN, MICHAEL B.
;    APPLICANT:  ORO, ANTHONY E.
;    APPLICANT:  SEGRAVES, WILLIAM A.
;    APPLICANT:  YAO, TSO-PANG
;    TITLE OF INVENTION:  MULTIMERIC FORMS OF MEMBERS OF THE
;    TITLE OF INVENTION:  STEROID/THYROID SUPERFAMILY OF RECEPTORS WITH THE
;    TITLE OF INVENTION:  ULTRASPIRACLE RECEPTOR
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PRETTY, SCHROEDER, BRUEGGEMANN & CLARK
;      STREET:  444 South Flower Street, Suite 2000
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/464,514
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/907,908
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter,, Stephen E.
;      REGISTRATION NUMBER:  31192
;      REFERENCE/DOCKET NUMBER:  P41 9321
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 546-4737
;      TELEFAX:  (619) 546-9392
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2304 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  163..1701
US-08-464-514-1

Alignment Scores:
Pred. No.:              0.00871       Length:          2304
Score:                  141.00        Matches:         106
Percent Similarity:     32.43%        Conservative:    49
Best Local Similarity:  22.18%        Mismatches:      127
Query Match:            7.38%         Indels:          197
DB:                     3             Gaps:            22

US-09-972-758A-2 (1-359) x US-08-464-514-1 (1-2304)

Qy        13 ProGlnThrSerAsnCysThrGlyAlaAlaAlaValGlnGluGluLeuAsnProGluArg 32
                |||   ::::::      :::   |||||||||||||||:::::      |||
Db       388 CCCAATTCGGCTGGAGGCTCTGCCGCCGCTGCAGTCCAGCAGCAGTATCCGCCTAACCAT 447

Qy        33 ProProGlyAlaGluGluArg---------ValProGluGluAspSerArgTrpGlnSer 49
                |||          :::          :::             :::|||||:::
Db       448 CCGCTGAGCGGCAGCAAGCACCTCTGCTCTATTTG-CGGGGATCGGGCCAGTGGCAAGCA 506

Qy        50 ArgAlaPheProGlnLeuGlyGlyArgProGly-----------ProGluGlyGluGly 65
                       ||||||   |||      |||                    :::|||
Db       507 CTACGGCGTGTACAGCTGTGAGGGCTGCAAGGGCTTCTTTAAACGCACAGTGCGCAAGGA 566

Qy        66 Ser---------LeuGluSerGlnProProProLeuGlnThrGlnAlaCysProGluSer 82
                |||         |||:::   :::|||               ||||||      |||
Db       567 TCTCACATACGCTTGCAGGGAGAACCGCAACTGCATCATAGACAAGCGGCAGAGGAA--- 623
```

```
Qy        83 SerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGlyGlyAsp 102

Db       623 ------------------------------------------------------------ 623

Qy       103 PheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGlnProCys 122
                 |||      |||   |||   ||||||         |||
Db       624 ---------CCGCTGCCAGTACTGCCGCTACCAGAAGTGCCTAACCTGCGG--------- 665

Qy       123 HisAspSerGluAlaSerLysLeuGlyAla----------ProAlaAlaGlyGlyGlu 138
             |||::::::   :::         ||||||            |||      ||||||
Db       666 CATGAAGCGCGAAGCGGTCCAGGAGGAGCGTCAACGCGGCGCCCGCAATGCGGCGGG--- 722

Qy       139 GluGluTrpGlyGlnGlnGlnArg---------------------------------- 146
                      |||:::||||||
Db       723 ------TAGGCTCAGCGCCAGCGGAGGCGGCAGTAGCGGTCCAGGTTCGGTAGGCGGATC 776

Qy       147 GlnLeuGlyLysLysLysHisArgArgArgProSerLysLysLysArgHisTrpLysPro 166
             ||||||   :::::::::   |||||||||                 |||   |||:::
Db       777 CAGCTCTCAAGGCGGAGGAGGAGGAGGCGG--------------CGTTTCTGGCGGAAT 821

Qy       167 TyrTyrLysLeuThrTrpGluGluLysLysLysPhe-AspGluLysGlnSerLeuArgAl 186
                           :::::::::::|||  |||:::    :::
Db       822 -----------------GGGCAGCGGCAACGGTTCTGATGACTTCATGACCAATAGCGT 863

Qy       186 aSerArg--------------------IleArgAlaGluMetPheAlaLysGlyGln-- 198
             ||||||                    |||   ||||||      |||:::   |||
Db       864 GTCCAGGGATTTCTCGATCGAGCGCATCATAGAGGCCGAGCAGCGAGCGGAGACCCAATG 923

Qy       199 ------------------------ProValAlaProTyrAsnThrThrGlnPheLeuMe 210
                                      |||   ||||||:::|||   |||
Db       924 CGGCGATCGTGCACTGACGTTCCTGCGCGTTGGTCCCTATTCCACAGTCCAG-------- 975

Qy       210 tAspAspHisAspGlnGluGluProAspLeuLysThrGlyLeu----------------- 224
                                   ||||||   |||      :::
Db       976 --------------------CCGGACTACAAGGGTGCCGTGTCGGCCCTGTGCCAAGT 1013

Qy       225 ------------------------------TyrSerLysArgAlaAlaAlaLysSerAs 234
                                           |||::::::                 :::
Db      1014 GGTCAACAAACAGCTCTTCCAGATGGTCGAATACGCGCGCATGATGCCGCACTTTGCCCA 1073

Qy       234 pAspThrSerAspAspAspPhe---------------------------------- 241
                 ||||||
Db      1074 GGTGCCGCTGGACGACCAGGTGATTCTGCTGAAAGCCGCTTGGATCGAGCTGCTCATTGC 1133

Qy       242 --------------------------MetGluGluGlyGlyGluGluAspGlyGlySe 252
                                       :::::::::||||||         ||||||
Db      1134 GAACGTGGCCTGGTGCAGCATCGTTTCGCTGGATGACGGCGGT---GCCGGCGGCGGGGG 1190

Qy       252 rAspGlyMetGlyGlyAspGlySer---------------------------------- 260
                ||:::|||   |||||||||
Db      1191 CGGTGGACTAGGCCACGATGGCTCCTTTGAGCGACGATCACCGGGCCTTCAGCCCCAGCA 1250

Qy       261 -GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeu-- 279
                ||||||   :::   ||||||              ||||||       |||
Db      1251 GCTGTTCCTCAACCAGAGCTTCTCG--------------TACCATCGCAACAGTGCGAT 1295

Qy       279 ------------------------------------------------------------ 279

Db      1296 CAAAGCCGGTGTGTCAGCCATCTTCGACCGCATATTGTCGGAGCTGAGTGTAAAGATGAA 1355

Qy       280 ----GlnAsnMetSerLysGlnGluLeu------------------------------ 287
                 |||:::   ::::::||||||
Db      1356 GCGGCTGAATCTCGACCGACGCGAGCTGTCCTGCTTGAAGGCCATCATACTGTACAACCC 1415

Qy       288 -------------IleLysGluTyrLeuGluLeuGluLysCysLeuSerArgMetGlu-- 302
                         ||||||         |||:::|||   |||      ::::::
Db      1416 GGACATACGCGGGATCAAGAGCCGGGCGGAGATCGAGATGTGCCGCGAGAAGGTGTACGC 1475

Qy       303 -------AspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAl 320
                 ||||||:::   ||||||                        |||||||||
Db      1476 TTGCCTGGACGAGCACTGCCGCCTGGAACATCCG--------------GGCGACGATGG 1520

Qy       320 aArgValArgGluLeuGluLeuGluLeuAspArg-LeuArgAlaGluAsnLeuGlnLeuL 340
```

```
                  |||      :::|||   |||   |||   ||| ||||||:::   :::|||:::
Db      1521 ACGCTTTGCGCAACTGCTGCTGCGTCTGCGCCGCTTTGCGATCGATCAGCCTGAAGTGCC 1580

Qy       340 euThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLys 356
                           |||   |||                ||||||   :::
Db      1581 AGGATCACCTGTTCCTCTTCCGCATTACCAGCGACCGGCCGCTGGAGGAG 1630

RESULT 12
US-08-486-403-1
; Sequence 1, Application US/08486403
; Patent No. 6281330
;  GENERAL INFORMATION:
;    APPLICANT:  EVANS, RONALD M.
;    APPLICANT:  MCKEOWN, MICHAEL B.
;    APPLICANT:  ORO, ANTHONY E.
;    APPLICANT:  SEGRAVES, WILLIAM A.
;    APPLICANT:  YAO, TSO-PANG
;    TITLE OF INVENTION:  MULTIMERIC FORMS OF MEMBERS OF THE
;    TITLE OF INVENTION:  STEROID/THYROID SUPERFAMILY OF RECEPTORS WITH THE
;    TITLE OF INVENTION:  ULTRASPIRACLE RECEPTOR
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PRETTY, SCHROEDER, BRUEGGEMANN & CLARK
;      STREET:  444 South Flower Street, Suite 2000
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/486,403
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/907,908
;      FILING DATE:  02-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter,, Stephen E.
;      REGISTRATION NUMBER:  31192
;      REFERENCE/DOCKET NUMBER:  P41 9321
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 546-4737
;      TELEFAX:  (619) 546-9392
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2304 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  163..1701
US-08-486-403-1

Alignment Scores:
Pred. No.:                0.00871        Length:          2304
Score:                    141.00         Matches:         106
Percent Similarity:       32.43%         Conservative:    49
Best Local Similarity:    22.18%         Mismatches:      127
Query Match:              7.38%          Indels:          197
DB:                       3              Gaps:            22

US-09-972-758A-2 (1-359) x US-08-486-403-1 (1-2304)

Qy         13 ProGlnThrSerAsnCysThrGlyAlaAlaAlaValGlnGluGluLeuAsnProGluArg 32
                 |||   ::::::      :::   ||||||||||||||||::::::      |||
Db        388 CCCAATTCGGCTGGAGGCTCTGCCGCCGCTGCAGTCCAGCAGCAGTATCCGCCTAACCAT 447
```

```
Qy         33 ProProGlyAlaGluGluArg---------ValProGluGluAspSerArgTrpGlnSer 49
Db        448 CCGCTGAGCGGCAGCAAGCACCTCTGCTCTATTTG-CGGGGATCGGGCCAGTGGCAAGCA 506

Qy         50 ArgAlaPheProGlnLeuGlyGlyArgProGly------------ProGluGlyGluGly 65
Db        507 CTACGGCGTGTACAGCTGTGAGGGCTGCAAGGGCTTCTTTAAACGCACAGTGCGCAAGGA 566

Qy         66 Ser---------LeuGluSerGlnProProProLeuGlnThrGlnAlaCysProGluSer 82
Db        567 TCTCACATACGCTTGCAGGGAGAACCGCAACTGCATCATAGACAAGCGGCAGAGGAA--- 623

Qy         83 SerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGlyGlyAsp 102
Db        623 ------------------------------------------------------------ 623

Qy        103 PheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGlnProCys 122
Db        624 ---------CCGCTGCCAGTACTGCCGCTACCAGAAGTGCCTAACCTGCGG--------- 665

Qy        123 HisAspSerGluAlaSerLysLeuGlyAla------------ProAlaAlaGlyGlyGlu 138
Db        666 CATGAAGCGCGAAGCGGTCCAGGAGGAGCGTCAACGCGCCGCCCGCAATGCGGCGGG--- 722

Qy        139 GluGluTrpGlyGlnGlnGlnArg---------------------------------- 146
Db        723 ------TAGGCTCAGCGCCAGCGGAGGCGGCAGTAGCGGTCCAGGTTCGGTAGGCGGATC 776

Qy        147 GlnLeuGlyLysLysLysHisArgArgArgProSerLysLysLysArgHisTrpLysPro 166
Db        777 CAGCTCTCAAGGCGGAGGAGGAGGAGGCGG--------------CGTTTCTGGCGGAAT 821

Qy        167 TyrTyrLysLeuThrTrpGluGluLysLysLysPhe-AspGluLysGlnSerLeuArgAl 186
Db        822 ------------------GGGCAGCGGCAACGGTTCTGATGACTTCATGACCAATAGCGT 863

Qy        186 aSerArg---------------------IleArgAlaGluMetPheAlaLysGlyGln-- 198
Db        864 GTCCAGGGATTTCTCGATCGAGCGCATCATAGAGGCCGAGCAGCGAGCGGAGACCCAATG 923

Qy        199 ------------------------ProValAlaProTyrAsnThrThrGlnPheLeuMe 210
Db        924 CGGCGATCGTGCACTGACGTTCCTGCGCGTTGGTCCCTATTCCACAGTCCAG-------- 975

Qy        210 tAspAspHisAspGlnGluGluProAspLeuLysThrGlyLeu----------------- 224
Db        976 ---------------------CCGGACTACAAGGGTGCCGTGTCGGCCCTGTGCCAAGT 1013

Qy        225 -----------------------------TyrSerLysArgAlaAlaAlaLysSerAs 234
Db       1014 GGTCAACAAACAGCTCTTCCAGATGGTCGAATACGCGCGCATGATGCCGCACTTTGCCCA 1073

Qy        234 pAspThrSerAspAspAspPhe---------------------------------- 241
Db       1074 GGTGCCGCTGGACGACCAGGTGATTCTGCTGAAAGCCGCTTGGATCGAGCTGCTCATTGC 1133

Qy        242 ---------------------------MetGluGluGlyGlyGluGluAspGlyGlySe 252
Db       1134 GAACGTGGCCTGGTGCAGCATCGTTTCGCTGGATGACGGCGGT---GCCGGCGGCGGGGG 1190

Qy        252 rAspGlyMetGlyGlyAspGlySer---------------------------------- 260
Db       1191 CGGTGGACTAGGCCACGATGGCTCCTTTGAGCGACGATCACCGGGCCTTCAGCCCCAGCA 1250

Qy        261 -GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeu-- 279
Db       1251 GCTGTTCCTCAACCAGAGCTTCTCG--------------TACCATCGCAACAGTGCGAT 1295

Qy        279 ------------------------------------------------------------ 279
Db       1296 CAAAGCCGGTGTGTCAGCCATCTTCGACCGCATATTGTCGGAGCTGAGTGTAAAGATGAA 1355

Qy        280 ----GlnAsnMetSerLysGlnGluLeu------------------------------- 287
Db       1356 GCGGCTGAATCTCGACCGACGCGAGCTGTCCTGCTTGAAGGCCATCATACTGTACAACCC 1415

Qy        288 -------------IleLysGluTyrLeuGluLeuGluLysCysLeuSerArgMetGlu-- 302
Db       1416 GGACATACGCGGGATCAAGAGCCGGGCGGAGATCGAGATGTGCCGCGAGAAGGTGTACGC 1475

Qy        303 -------AspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAl 320
Db       1476 TTGCCTGGACGAGCACTGCCGCCTGGAACATCCG--------------GGCGACGATGG 1520

Qy        320 aArgValArgGluLeuGluLeuGluLeuAspArg-LeuArgAlaGluAsnLeuGlnLeuL 340
Db       1521 ACGCTTTGCGCAACTGCTGCTGCGTCTGCGCCGCTTTGCGATCGATCAGCCTGAAGTGCC 1580

Qy        340 euThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLys 356
Db       1581 AGGATCACCTGTTCCTCTTCCGCATTACCAGCGACCGGCCGCTGGAGGAG 1630
```

RESULT 13
US-09-265-013-2
; Sequence 2, Application US/09265013
; Patent No. 6451304
; GENERAL INFORMATION:
; APPLICANT: FRIEDMAN, Theodore
; APPLICANT: MIYANOHARA, Atsushi
; TITLE OF INVENTION: METHOD FOR RETROVIRUS VECTOR PRODUCTION BY SEPARATED
; TITLE OF INVENTION: GAG AND POL EXPRESSION
; FILE REFERENCE: 041673/2010
; CURRENT APPLICATION NUMBER: US/09/265,013
; CURRENT FILING DATE: 1999-03-09
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1617
; TYPE: DNA
; ORGANISM: Moloney murine leukemia virus (MoMLV)
US-09-265-013-2

Alignment Scores:

Pred. No.:	0.0117	Length:	1617
Score:	137.00	Matches:	90
Percent Similarity:	30.63%	Conservative:	42
Best Local Similarity:	20.88%	Mismatches:	152
Query Match:	7.17%	Indels:	147
DB:	4	Gaps:	17

US-09-972-758A-2 (1-359) x US-09-265-013-2 (1-1617)

```
Qy         18 CysThrGly---AlaAlaAlaValGlnGluGluLeuAsnProGluArgProPro-GlyAl 36
Db        300 TGTACACCCTAAGCCTCCGCCTCCTCTTCCTCCATCCGCCCCGTCTCTCCCCCTTGAACC 359

Qy         36 aGluGluArgValProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGl 56
Db        360 TCCTCGTTCGACCCCGCCTCGATCCTCCCTTTATCCAGCCCTCACT---CCTTCTCTAGG 416

Qy         56 yGlyArgProGlyProGlu---------GlyGluGlySer---------------LeuGl 68
Db        417 CGCCAAACCTAAACCTCAAGTTCTTTCTGACAGTGGGGGGCCGCTCATCGACCTACTTAC 476

Qy         68 uSerGlnProProProLeuGlnThrGlnAlaCysProGluSerSerCysLeuArgGluGl 88
Db        477 AGAAGACCCCCCGCCTTATAGGGACCCAAGACCACCCCCTTCCGAC------AGGGACGG 530

Qy         88 yGluLysGlyGlnAsnGlyAspAspSerSerAlaGlyGlyAspPheProProProAlaGl 108
Db        531 A------------AATGGTGGAGAAGCGACCCCTGCGGGAGAGGCACCG---------- 567

Qy        108 uValGluProThrProGluAlaGluLeuLeu------------AlaGlnProCysHisAs 124
Db        568 ----GACCCCTCCCCAATGGCATCTCGCCTACGTGGGAGACGGGAGCCCCCTGTGGCCGA 623
```

```
Qy       124 pSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGluTrpGlyGlnGl 144
             ||||    |||:::           ||||||||
Db       624 CTCCACTACCTCGCAGGCATTCCCCCTCCGCGCAGGAGGAAAC---------------- 666

Qy       144 nGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLysLysArgHisTr 164
                     |||:::   ::::::           |||           :::||
Db       667 ------------GGACAGCTTCAATACTGGCCGTTCTCCTCTTCTGACCTTTACAACTG 713

Qy       164 pLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGluLysGlnSerLe 184
             ||||                                     |||   |||
Db       714 GAAA--------------------AATAATAACCCTTCTTTTTCTGAAGATCCAGGTAA 752

Qy       184 uArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProValAlaProTyrAs 204
                 :::   |||   :::   :::
Db       753 ACTGACAGCTCTGATCGAGTCTGTTCTCATC---------------------------- 783

Qy       204 nThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeuLysThrGly-- 223
              |||   |||         ||||||      |||           |||   ||||||
Db       784 -ACCCATCAGCCCACCTGGGACGACTGTCAGCAGCTGTTGGGGACTCTGCTGACCGGAGA 842

Qy       224 ---------------LeuTyrSerLysArgAlaAlaAlaLysSerAspAsp-------- 235
                             |||          |||   |||   :::   ||||||
Db       843 AGAAAAACAACGGGTGCTCTTAGAGGCTAGAAAGGCGGTGCGGGGCGATGATGGGCGCCC 902

Qy       236 ----------------------------------------------ThrSerAspAspAspPh 241
                                                           |||   |||::
Db       903 CACTCAACTGCCCAATGAAGTCGATGCCGCTTTTCCCCTCGAGCGCCCAGACTGGGATTA 962

Qy       241 eMetGluGluGlyGlyGluGlu------------------------------------- 248
             :      :::   |||
Db       963 CACCACCCAGGCAGGTAGGAACCACCTAGTCCACTATCGCCAGTTGCTCCTAGCGGGTCT 1022

Qy       249 ----------------------------------------AspGlyGlySerAspGlyMe 255
                                                    |||   :::   |||
Db      1023 CCAAAACGCGGGCAGAAGCCCCACCAATTTGGCCAAGGTAAAAGGAATAACACAAGGGCC 1082

Qy       255 tGlyGlyAspGlySerGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyr-- 274
                 |||   ||||||:::|||         |||   |||   ||||||
Db      1083 CAATGAGTCTCCCTCGGCCTTCCTAGAGAGA---CTTAAGGAAGCCTATCGCAGGTACAC 1139

Qy       274 ------------------------------------------------------------ 274

Db      1140 TCCTTATGACCCTGAGGACCCAGGGCAAGAAACTAATGTGTCTATGTCTTTCATTTGGCA 1199

Qy       275 -------------------------------HisThrGluSerLeuGlnAsnMetSerLy 284
                                             |||   |||:::|||   :::
Db      1200 GTCTGCCCCAGACATTGGGAGAAAGTTAGAGAGGTTAGAAGATTTAAAAAACAAGACGCT 1259

Qy       284 sGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArgMetGluAspGl 304
                 :::|||::4:::|||      ::::::                ||||:::
Db      1260 TGGAGATTTGGTTAGAGAGGCAGAAAAGATCTTTAATAAACGAGAAACCCCGGAAGAAAG 1319

Qy       304 uAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAlaArgValArgGl 324
                   |||:::|||                                     |||||
Db      1320 AGAGGAACGTATCAGG-----------------------------------AGAGA 1340

Qy       324 uLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnGl 344
             |   |||   :::   :::|||   |||   |||:::   |||           |||   ::
Db      1341 AACAGAGGAAAAAGAAGAACGCCGTAGGACAGAGGATGAGCAGAAAGAGAAAGAAAGAGA 1400

Qy       344 uLeuHisArgGlnGlnGluArgAlaProLeu 354
             :      |||   :::|||   :::   |||
Db      1401 TCGTAGGAGACATAGAGAGATGAGCAAGCTA 1431

RESULT 14
US-09-011-745-3
; Sequence 3, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
;  APPLICANT: Collins, Mary KL
;  APPLICANT: Weiss, Robin A
;  APPLICANT: Takeuchi, Yasuhiro
;  APPLICANT: Cosset, Francois-Loic
;  TITLE OF INVENTION: Expression systems
;  FILE REFERENCE: 09/011,745
;  CURRENT APPLICATION NUMBER: US/09/011,745
;  CURRENT FILING DATE: 1998-06-22
;  EARLIER APPLICATION NUMBER: PCT/GB96/02061
;  EARLIER FILING DATE: 1996-08-23
;  EARLIER APPLICATION NUMBER: GB9517263.1
;  EARLIER FILING DATE: 1995-08-23
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 7308
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Portion of
;   OTHER INFORMATION: construct
US-09-011-745-3

Alignment Scores:
Pred. No.:                0.0938          Length:          7308
Score:                    137.00          Matches:         90
Percent Similarity:       30.63%          Conservative:    42
Best Local Similarity:    20.88%          Mismatches:      152
Query Match:              7.17%           Indels:          147
DB:                       3               Gaps:            17

US-09-972-758A-2 (1-359) x US-09-011-745-3 (1-7308)

Qy        18 CysThrGly---AlaAlaAlaValGlnGluGluLeuAsnProGluArgProPro-GlyAl 36
             ||||||      |||:::|||            :::   |||       ||||||
Db      1866 TGTACACCCTAAGCCTCCGCCTCCTCTTCCTCCATCCGCCCCGTCTCTCCCCCTTGAACC 1925

Qy        36 aGluGluArgValProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGl 56
                      |||       |||   :::   :::       |||   |||||
Db      1926 TCCTCGTTCGACCCCGCCTCGATCCTCCCTTTATCCAGCCCTCACT---CCTTCTCTAGG 1982

Qy        56 yGlyArgProGlyProGlu---------GlyGluGlySer--------------LeuGl 68
             |   :::|||   |||:::              |||                   |||
Db      1983 CGCCAAACCTAAACCTCAAGTTCTTTCTGACAGTGGGGGGCCGCTCATCGACCTACTTAC 2042

Qy        68 uSerGlnProProProLeuGlnThrGlnAlaCysProGluSerSerCysLeuArgGluGl 88
                  |||||||||   :::           |||   |||         |||:::||
Db      2043 AGAAGACCCCCCGCCTTATAGGGACCCAAGACCACCCCCTTCCGAC------AGGGACGG 2096

Qy        88 yGluLysGlyGlnAsnGlyAspAspSerSerAlaGlyGlyAspPheProProProAlaGl 108
             |          ||||||   :::::::::       |||:::   |||
Db      2097 A------------AATGGTGGAGAAGCGACCCCTGCGGGAGAGGCACCG----------- 2133

Qy       108 uValGluProThrProGluAlaGluLeuLeu------------AlaGlnProCysHisAs 124
                 :::|||:::|||   |||       |||                  |||       ||
Db      2134 ----GACCCCTCCCCAATGGCATCTCGCCTACGTGGGAGACGGGAGCCCCCTGTGGCCGA 2189

Qy       124 pSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGluTrpGlyGlnGl 144
             ||||      |||:::                |||||||||
Db      2190 CTCCACTACCTCGCAGGCATTCCCCCTCCGCGCAGGAGGAAAC---------------- 2232

Qy       144 nGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLysLysArgHisTr 164
                     |||:::   ::::::           |||           :::||
Db      2233 ------------GGACAGCTTCAATACTGGCCGTTCTCCTCTTCTGACCTTTACAACTG 2279

Qy       164 pLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGluLysGlnSerLe 184
             ||||                                     |||   |||
Db      2280 GAAA--------------------AATAATAACCCTTCTTTTTCTGAAGATCCAGGTAA 2318

Qy       184 uArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProValAlaProTyrAs 204
                 :::   |||   :::   :::
Db      2319 ACTGACAGCTCTGATCGAGTCTGTTCTCATC---------------------------- 2349

Qy       204 nThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeuLysThrGly-- 223
              |||   |||         ||||||      |||           |||   ||||||
Db      2350 -ACCCATCAGCCCACCTGGGACGACTGTCAGCAGCTGTTGGGGACTCTGCTGACCGGAGA 2408
```

```
Qy        224 ----------------LeuTyrSerLysArgAlaAlaAlaLysSerAspAsp-------- 235
Db       2409 AGAAAAACAACGGGTGCTCTTAGAGGCTAGAAAGGCGGTGCGGGGCGATGATGGGCGCCC 2468

Qy        236 ------------------------------------------ThrSerAspAspAspPh 241
Db       2469 CACTCAACTGCCCAATGAAGTCGATGCCGCTTTTCCCCTCGAGCGCCCAGACTGGGATTA 2528

Qy        241 eMetGluGluGlyGlyGluGlu----------------------------------- 248
Db       2529 CACCACCCAGGCAGGTAGGAACCACCTAGTCCACTATCGCCAGTTGCTCCTAGCGGGTCT 2588

Qy        249 ----------------------------------------AspGlyGlySerAspGlyMe 255
Db       2589 CCAAAACGCGGGCAGAAGCCCCACCAATTTGGCCAAGGTAAAAGGAATAACACAAGGGCC 2648

Qy        255 tGlyGlyAspGlySerGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyr-- 274
Db       2649 CAATGAGTCTCCCTCGGCCTTCCTAGAGAGA---CTTAAGGAAGCCTATCGCAGGTACAC 2705

Qy        274 ------------------------------------------------------------ 274
Db       2706 TCCTTATGACCCTGAGGACCCAGGGCAAGAAACTAATGTGTCTATGTCTTTCATTTGGCA 2765

Qy        275 ------------------------------HisThrGluSerLeuGlnAsnMetSerLy 284
Db       2766 GTCTGCCCCAGACATTGGGAGAAAGTTAGAGAGGTTAGAAGATTTAAAAAACAAGACGCT 2825

Qy        284 sGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArgMetGluAspGl 304
Db       2826 TGGAGATTTGGTTAGAGAGGCAGAAAAGATCTTTAATAAACGAGAAACCCCGGAAGAAAG 2885

Qy        304 uAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAlaArgValArgGl 324
Db       2886 AGAGGAACGTATCAGG-------------------------------------AGAGA 2906

Qy        324 uLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnGl 344
Db       2907 AACAGAGGAAAAAGAAGAACGCCGTAGGACAGAGGATGAGCAGAAAGAGAAAGAAAGAGA 2966

Qy        344 uLeuHisArgGlnGlnGluArgAlaProLeu 354
Db       2967 TCGTAGGAGACATAGAGAGATGAGCAAGCTA 2997
```

RESULT 15
US-09-011-745-4
; Sequence 4, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A
; APPLICANT: Takeuchi, Yasuhiro
; APPLICANT: Cosset, Francois-Loic
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; CURRENT FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; EARLIER FILING DATE: 1995-08-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 7308
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Portion of
; OTHER INFORMATION: construct
US-09-011-745-4

Alignment Scores:

Pred. No.:	0.0938	Length:	7308
Score:	137.00	Matches:	90
Percent Similarity:	30.63%	Conservative:	42
Best Local Similarity:	20.88%	Mismatches:	152
Query Match:	7.17%	Indels:	147
DB:	3	Gaps:	17

US-09-972-758A-2 (1-359) x US-09-011-745-4 (1-7308)

```
Qy         18 CysThrGly---AlaAlaAlaValGlnGluGluLeuAsnProGluArgProPro-GlyAl 36
Db       1866 TGTACACCCTAAGCCTCCGCCTCCTCTTCCTCCATCCGCCCCGTCTCTCCCCCTTGAACC 1925

Qy         36 aGluGluArgValProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGl 56
Db       1926 TCCTCGTTCGACCCCGCCTCGATCCTCCCTTTATCCAGCCCTCACT---CCTTCTCTAGG 1982

Qy         56 yGlyArgProGlyProGlu---------GlyGluGlySer---------------LeuGl 68
Db       1983 CGCCAAACCTAAACCTCAAGTTCTTTCTGACAGTGGGGGGCCGCTCATCGACCTACTTAC 2042

Qy         68 uSerGlnProProProLeuGlnThrGlnAlaCysProGluSerSerCysLeuArgGluGl 88
Db       2043 AGAAGACCCCCCGCCTTATAGGGACCCAAGACCACCCCCTTCCGAC------AGGGACGG 2096

Qy         88 yGluLysGlyGlnAsnGlyAspAspSerSerAlaGlyGlyAspPheProProProAlaGl 108
Db       2097 A------------AATGGTGGAGAAGCGACCCCTGCGGGAGAGGCACCG----------- 2133

Qy        108 uValGluProThrProGluAlaGluLeuLeu------------AlaGlnProCysHisAs 124
Db       2134 ----GACCCCTCCCCAATGGCATCTCGCCTACGTGGGAGACGGGAGCCCCCTGTGGCCGA 2189

Qy        124 pSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGluTrpGlyGlnGl 144
Db       2190 CTCCACTACCTCGCAGGCATTCCCCCTCCGCGCAGGAGGAAAC----------------- 2232

Qy        144 nGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLysLysArgHisTr 164
Db       2233 ------------GGACAGCTTCAATACTGGCCGTTCTCCTCTTCTGACCTTTACAACTG 2279

Qy        164 pLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGluLysGlnSerLe 184
Db       2280 GAAA--------------------AATAATAACCCTTCTTTTTCTGAAGATCCAGGTAA 2318

Qy        184 uArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProValAlaProTyrAs 204
Db       2319 ACTGACAGCTCTGATCGAGTCTGTTCTCATC---------------------------- 2349

Qy        204 nThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeuLysThrGly-- 223
Db       2350 -ACCCATCAGCCCACCTGGGACGACTGTCAGCAGCTGTTGGGGACTCTGCTGACCGGAGA 2408

Qy        224 ----------------LeuTyrSerLysArgAlaAlaAlaLysSerAspAsp-------- 235
Db       2409 AGAAAAACAACGGGTGCTCTTAGAGGCTAGAAAGGCGGTGCGGGGCGATGATGGGCGCCC 2468

Qy        236 ------------------------------------------ThrSerAspAspAspPh 241
Db       2469 CACTCAACTGCCCAATGAAGTCGATGCCGCTTTTCCCCTCGAGCGCCCAGACTGGGATTA 2528

Qy        241 eMetGluGluGlyGlyGluGlu----------------------------------- 248
Db       2529 CACCACCCAGGCAGGACGCAACCACCTAGTCCACTATCGCCAGTTGCTCCTAGCGGGTCT 2588

Qy        249 ----------------------------------------AspGlyGlySerAspGlyMe 255
Db       2589 CCAAAACGCGGGCAGAAGCCCCACCAATTTGGCCAAGGTAAAAGGAATAACACAAGGGCC 2648

Qy        255 tGlyGlyAspGlySerGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyr-- 274
Db       2649 CAATGAGTCTCCCTCGGCCTTCCTAGAGAGA---CTTAAGGAAGCCTATCGCAGGTACAC 2705
```

```
Qy        274 ------------------------------------------------------------ 274

Db       2706 TCCTTATGACCCTGAGGACCCAGGGCAAGAAACTAATGTGTCTATGTCTTTCATTTGGCA 2765

Qy        275 ------------------------------HisThrGluSerLeuGlnAsnMetSerLy 284
                                             |||   |||:::|||   :::
Db       2766 GTCTGCCCCAGACATTGGGAGAAAGTTAGAGAGGTTAGAAGATTTAAAAAACAAGACGCT 2825

Qy        284 sGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArgMetGluAspGl 304
                  :::|||::::::|||      ::::::                      |||:::
Db       2826 TGGAGATTTGGTTAGAGAGGCAGAAAAGATCTTTAATAAACGAGAAACCCCGGAAGAAAG 2885

Qy        304 uAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAlaArgValArgGl 324
                    |||:::|||                                      |||||
Db       2886 AGAGGAACGTATCAGG------------------------------------AGAGA 2906

Qy        324 uLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnGl 344
                  |  |||   :::   :::|||   |||   |||:::   |||         |||   ::
Db       2907 AACAGAGGAAAAAGAAGAACGCCGTAGGACAGAGGATGAGCAGAAAGAGAAAGAAAGAGA 2966

Qy        344 uLeuHisArgGlnGlnGluArgAlaProLeu 354
                :      |||   :::|||   :::   |||
Db       2967 TCGTAGGAGACATAGAGAGATGAGCAAGCTA 2997
```

Search completed: February 5, 2004, 15:52:30
Job time : 124 secs

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 5, 2004, 13:32:05 ; Search time 395 Seconds
(without alignments)
2453.413 Million cell updates/sec

Title: US-09-972-758A-2
Perfect score: 1910
Sequence: 1 MAEPFLSEYQHQPQTSNCTG.....LTENELHRQQERAPLSKFGD 359

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09972758/runat_05022004_095002_1181/app_query.fasta_1.519
-DB=N_Geneseq_19Jun03 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09972758_@CGN_1_1_0_@runat_05022004_095002_1181 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1910	100.0	2199	19	AAV82778	Clone bp783_3 isol
2		1910	100.0	2199	24	ABQ92015	Human polynucleoti
3		1903	99.6	2178	22	AAH18098	Human cDNA sequenc
4		1902	99.6	1080	24	ABN84013	Human oestrogen do
5	c	733	38.4	1083	24	ABL90067	Human polynucleoti
6		714	37.4	523	20	AAV86036	EST clone B227. H
7		691	36.2	414	25	ABX54966	Bovine EST associa
8		629	32.9	461	24	ABL63543	Breast cancer rela
9		545.5	28.6	1557	24	ABQ61099	Mouse putative pro
10	c	308.5	16.2	997	22	AAH99667	Human protein enco
11		245	12.8	342	20	AAV87316	EST clone BP783.
12		214	11.2	1489	23	ABL06283	Drosophila melanog
13	c	205.5	10.8	3714	23	ABL06282	Drosophila melanog
14		184	9.6	549	21	AAC95217	Cat flea head and
15		174	9.1	1024	23	AAS79493	DNA encoding novel
16		162	8.5	7396	24	ABT08505	Human novel protei
17		162	8.5	7516	22	AAK51987	Human polynucleoti
18		159.5	8.4	4041	24	ABQ54663	Human ovarian anti
19		159.5	8.4	7716	23	AAS65910	DNA encoding novel
20		158.5	8.3	9551	20	AAZ22301	cDNA encoding a hu
21		156	8.2	6432	23	AAS65903	DNA encoding novel
22		155	8.1	2277	19	AAV13836	Homo sapiens mamma
23		155	8.1	2277	19	AAV05372	Human telomerase p
24		153.5	8.0	3579	23	ABL05651	Drosophila melanog
25	c	153.5	8.0	6105	23	ABL05650	Drosophila melanog
26		152	8.0	2421	24	ABK80600	Bacillus clausii g
27		152	8.0	3540	23	ABL06305	Drosophila melanog
28	c	151	7.9	3707	23	ABL06276	Drosophila melanog
29		149.5	7.8	8486	22	AAK52971	Human polynucleoti
30		149	7.8	2483	23	ABL07559	Drosophila melanog
31	c	149	7.8	4483	23	ABL07558	Drosophila melanog
32	c	149	7.8	18737	23	ABL07492	Drosophila melanog
33	c	149	7.8	18737	23	ABL07530	Drosophila melanog
34		148.5	7.8	7741	22	AAI60155	Human polynucleoti
35		148.5	7.8	7741	22	AAI60156	Human polynucleoti
36		148	7.7	5943	22	AAS22920	DNA encoding novel
37		147.5	7.7	3773	21	AAF18028	Lung cancer associ
38		147.5	7.7	15606	23	ABL16881	Drosophila melanog
39		147	7.7	7453	22	AAI58369	Human polynucleoti
40		147	7.7	7501	22	AAI58370	Human polynucleoti
41		145	7.6	4301	24	ABK63759	Rat sequence diffe
42	c	145	7.6	8731	23	ABL10296	Drosophila melanog
43		144	7.5	4714	22	AAD21393	Human Cockayne syn
44		143.5	7.5	2159	15	AAQ67223	Mouse p55Nuc. Mus
45		143	7.5	867	23	AAS84549	DNA encoding novel

ALIGNMENTS

RESULT 1
AAV82778
ID AAV82778 standard; cDNA; 2199 BP.
XX
AC AAV82778;
XX
DT 25-FEB-1999 (first entry)
XX
DE Clone bp783_3 isolated from human foetal kidney cDNA library.
XX
KW Secreted protein; nutritional activity; immune stimulating; vaccine;
KW suppressing activity; haematopoiesis regulating activity;
KW tissue growth activity; activin; inhibin activity; chemotactaxis;
KW chemokinetic activity; haemostasis; thrombolytic activity; receptor;
KW ligand; anti-inflammatory; cadherin; tumour invasion suppressor;
KW tumour inhibition; gene therapy; ds.
XX

OS Homo sapiens.
XX
PN WO9842739-A2.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US05653.
XX
PR 19-MAR-1998; 98US-0044466.
PR 21-MAR-1997; 97US-0822167.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX
DR WPI; 1998-609890/51.
DR P-PSDB; AAW85455.
XX
PT New polynucleotides encoding secreted human proteins - derived from
PT human foetal brain, adult brain, foetal kidney, placenta or adult
PT pineal gland cDNA libraries.
XX
PS Claim 1; Page 66-67; 113pp; English.
XX
CC The present sequence encodes a secreted protein. The polynucleotide and
CC secreted protein are predicted to have biological activities which would
CC make them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is given.
CC Suggested activities include nutritional activity, immune stimulating
CC (e.g. as vaccines) or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC invasion suppressor activity, and tumour inhibition activity (no data is
CC given in the specification to support these activities). The
CC polynucleotide is also stated to be useful for gene therapy.
XX
SQ Sequence 2199 BP; 552 A; 511 C; 674 G; 462 T; 0 other;

Alignment Scores:

Pred. No.:	8.18e-106	Length:	2199
Score:	1910.00	Matches:	359
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	19	Gaps:	0

US-09-972-758A-2 (1-359) x AAV82778 (1-2199)

```
Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        707 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 766

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        767 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCCAGGCGCGGAGGAGCGGGTG 826

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        827 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 886

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        887 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 946

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        947 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 1006

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1007 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 1066

Qy        121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1067 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 1126

Qy        141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1127 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 1186

Qy        161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1187 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 1246

Qy        181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1247 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 1306

Qy        201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1307 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 1366

Qy        221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1367 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 1426

Qy        241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1427 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 1486

Qy        261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 1546

Qy        281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArg 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 1606

Qy        301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 1666

Qy        321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1726

Qy        341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGAC 1783
```

RESULT 2
ABQ92015
ID ABQ92015 standard; cDNA; 2199 BP.
XX
AC ABQ92015;
XX
DT 04-OCT-2002 (first entry)
XX
DE Human polynucleotide SEQ ID NO 12.
XX
KW Human; cytostatic; antirheumatic; antiarthritic; vulnerary; analgesic;
KW antiinflammatory; antibacterial; immunosuppressive; antiparkinsonian;
KW neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;
KW antiulcer; fungicide; antidiabetic; antiasthmatic; antiallergic;
KW immunostimulant; antiparasitic; secreted protein; transmembrane protein;
KW cytokine; cell proliferation; cell differentiation; autoimmune disease;
KW stem cell; growth factor; nervous system disease; neuropathy;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW osteoporosis; severe combined immunodeficiency; SCID; infection;
KW multiple sclerosis; rheumatoid arthritis; gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN US2002065394-A1.

XX
PD 30-MAY-2002.
XX
PF 22-DEC-2000; 2000US-0745763.
XX
PR 18-MAR-1998; 98US-0040963.
XX
PA (JACO/) JACOBS K.
PA (MCCO/) MCCOY J M.
PA (LAVA/) LAVALLIE E R.
PA (COLL/) COLLINS-RACIE L A.
PA (EVAN/) EVANS C.
PA (MERB/) MERBERG D.
PA (TREA/) TREACY M.
PA (SPAU/) SPAULDING V.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Spaulding V;
XX
DR WPI; 2002-582343/62.
DR P-PSDB; ABP61799.
XX
PT Novel secreted or transmembrane protein and polynucleotide encoding the
PT protein, useful for diagnosis and treatment of neurological disorders,
PT cancer, autoimmune diseases, bone disorders and lung or liver fibrosis
PT -
XX
PS Claim 37; Page 111-112; 284pp; English.
XX
CC The invention relates to human secreted or transmembrane protein (I),
CC their fragments and is encoded by specific complementary deoxyribonucleic
CC acid (cDNA) inserts (II), where the protein is substantially free from
CC other mammalian proteins. (I) are useful for preventing, treating or
CC ameliorating a medical condition, especially immunological treatment or
CC prevention of tumours. (I) exhibits activity relating to angiogenesis,
CC cytokine, cell proliferation, cell differentiation, antiinflammatory,
CC stem cell growth factor activity and activin or inhibin-related
CC activities. (I) can be used to manipulate stem cells in culture to give
CC rise to neuroepithelial cells that can be used to augment or replace
CC cells damaged by illness, autoimmune disease, accidental damage or
CC genetic disorders. (I) induces the proliferation of neural cells and
CC regeneration of nerve and brain tissue and is useful for the treatment of
CC central and peripheral nervous system diseases and neuropathies, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis. (I) is involved in chemotactic or chemokinetic
CC activity, regulation of haematopoiesis and is useful for treating myeloid
CC or lymphoid cell disorders, platelet disorders such as thrombocytopaenia
CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve
CC tissue growth and in tissue repair, healing of burns, incisions, ulcers,
CC for treating osteoporosis, osteoarthritis, bone degenerative disorders or
CC periodontal disease. (I) is also useful for gut protection or
CC regeneration and treatment of lung or liver fibrosis, reperfusion injury
CC in various tissues, various immune deficiencies and disorders including
CC severe combined immunodeficiency (SCID), bacterial or fungal infections,
CC autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis,
CC diabetes mellitus, myasthenia gravis, allergic reactions and conditions,
CC such as asthma or other respiratory problems. (II) is useful to express
CC recombinant protein, as markers for tissues in which the corresponding
CC protein is preferentially expressed and in gene therapy. The present
CC sequence is that of a polynucleotide of the invention.
XX
SQ Sequence 2199 BP; 552 A; 511 C; 674 G; 462 T; 0 other;

Alignment Scores:

Pred. No.:	8.18e-106	Length:	2199
Score:	1910.00	Matches:	359
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	24	Gaps:	0

US-09-972-758A-2 (1-359) x ABQ92015 (1-2199)

```
Qy        1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
Db      707 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 766

Qy       21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
Db      767 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 826

Qy       41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
Db      827 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 886

Qy       61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
Db      887 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 946

Qy       81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
Db      947 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 1006

Qy      101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
Db     1007 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 1066

Qy      121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140
Db     1067 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 1126

Qy      141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
Db     1127 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 1186

Qy      161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
Db     1187 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 1246

Qy      181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
Db     1247 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 1306

Qy      201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
Db     1307 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 1366

Qy      221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240
Db     1367 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 1426

Qy      241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260
Db     1427 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 1486

Qy      261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280
Db     1487 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 1546

Qy      281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArg 300
Db     1547 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 1606

Qy      301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320
Db     1607 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 1666

Qy      321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340
Db     1667 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1726

Qy      341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359
Db     1727 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGAC 1783
```

RESULT 3

```
AAH18098
ID   AAH18098 standard; cDNA; 2178 BP.
XX
AC   AAH18098;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:17952.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 17952; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 2178 BP; 525 A; 513 C; 678 G; 462 T; 0 other;

Alignment Scores:
Pred. No.:                2.13e-105      Length:          2178
Score:                    1903.00        Matches:         358
Percent Similarity:       99.72%         Conservative:    0
Best Local Similarity:    99.72%         Mismatches:      1
Query Match:              99.63%         Indels:          0
DB:                       22             Gaps:            0

US-09-972-758A-2 (1-359) x AAH18098 (1-2178)

Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        716 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 775

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        776 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 835

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        836 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 895

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        896 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 955

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        956 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 1015

Qy        101 GlyAspProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1016 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 1075

Qy        121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1076 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 1135

Qy        141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1136 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 1195

Qy        161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1196 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 1255

Qy        181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1256 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 1315

Qy        201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1316 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 1375

Qy        221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1376 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 1435

Qy        241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1436 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 1495

Qy        261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1496 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 1555

Qy        281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArg 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1556 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 1615

Qy        301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1616 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 1675

Qy        321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1676 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1735

Qy        341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359
              ||||||||||||||||||||||||||||||||||||  ||||||||||||||
Db       1736 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCCTTCCAAGTTTGGAGAC 1792
```

```
RESULT 4
ABN84013
ID   ABN84013 standard; cDNA; 1080 BP.
XX
AC   ABN84013;
XX
DT   10-SEP-2002  (first entry)
XX
DE   Human oestrogen downregulated gene EDG1 coding sequence.
XX
KW   EDG1; oestrogen downregulated gene; tumour suppressor; human;
KW   breast cancer; prostate cancer; testicular cancer; ovarian cancer;
KW   uterine cancer; colon cancer; chromosome 17q; gene therapy; gene;
KW   ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1080
FT                   /*tag=  a
FT                   /product= "EDG1"
FT                   /transl_except= "(pos:460..462,aa:Arg)"
FT                   /transl_except= "(pos:511..513,aa:Thr)"
XX
PN   WO200228879-A1.
XX
PD   11-APR-2002.
XX
PF   05-OCT-2001; 2001WO-US31300.
XX
PR   05-OCT-2000; 2000US-238187P.
XX
PA   (UYCA-) UNIV CASE WESTERN RESERVE.
XX
PI   Montano M,  Wittman B;
XX
DR   WPI; 2002-519107/55.
DR   P-PSDB; ABB76495.
XX
PT   Polynucleotides encoding Estrogen Down-Regulated Gene 1 proteins,
PT   useful for the prevention, diagnosis and treatment of e.g. breast
PT   cancer, testicular cancer, prostate cancer, uterine cancer, cervical
PT   cancer, ovarian cancer and colon cancer -
XX
PS   Claim 1; Fig 1A-B; 52pp; English.
XX
CC   The present sequence is the coding sequence for human oestrogen
CC   downregulated gene 1 (EDG1), a tumour suppressor gene that is
CC   downregulated by oestrogen in mammary epithelial cells.  The gene
CC   was identified by yeast two-hybrid screenings for oestrogen
CC   receptor-interacting proteins in breast epithelial cells.  It
CC   was localised to chromosome arm 17q.  EDG1 mRNA expression is
CC   prevalent in normal mammary epithelial cells and in other human
CC   hormone-responsive tissues such as the ovary, prostate and testis.
CC   Expression is low in breast cancer epithelial cells.  Oestradiol,
CC   which induces breast cancer cell growth, has an inhibitory effect
CC   on EDG1 mRNA expression in breast cancer cells.  Hexamethylene
CC   bis-acetamide, an inducer of differentiation and apoptosis,
CC   upregulates EDG1 mRNA expression in breast cancer cells.  The
CC   invention provides EDG1 polynucleotides and polypeptides.  In a
CC   claimed method, a test sample from an individual suspected of
CC   having, or known to have breast, testicular, prostate, uterine,
CC   cervical, ovarian or colon cancer is assayed for EDG1 transcript
CC   using a polynucleotide that is complementary to the present
CC   sequence or by RT-PCR using a primer derived from the present
CC   sequence.  A decrease in the level of transcript compared to the
CC   level in a test sample indicates that the test sample contains or
CC   was derived from cancerous cells antibody.  A claimed method for
CC   decreasing the proliferation of breast, prostate, testicular,
CC   ovarian, uterine, cervical or colon cancer cells involves increasing
CC   EDG1 protein activity in the cells, either by contacting the cells
CC   with EDG1 protein or its fragment or functional equivalent, or with
CC   a nucleic acid encoding EDG1 protein, its fragment or functional
CC   equivalent.
XX
SQ   Sequence 1080 BP; 265 A; 296 C; 376 G; 143 T; 0 other;

Alignment Scores:
Pred. No.:              1.19e-105      Length:          1080
Score:                  1902.00        Matches:         357
Percent Similarity:     99.72%         Conservative:    1
Best Local Similarity:  99.44%         Mismatches:      1
Query Match:            99.58%         Indels:          0
DB:                     24             Gaps:            0

US-09-972-758A-2 (1-359) x ABN84013 (1-1080)

Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360

Qy        121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 420

Qy        141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        421 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAG 480

Qy        161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
              |||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db        481 AAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAG 540

Qy        181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 600

Qy        201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 660

Qy        221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 720

Qy        241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 780

Qy        261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 840

Qy        281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArg 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 900
```

```
Qy        301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 960

Qy        321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1020

Qy        341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGAC 1077
```

RESULT 5
ABL90067/c
ID ABL90067 standard; cDNA; 1083 BP.
XX
AC ABL90067;
XX
DT 24-MAY-2002 (first entry)
XX
DE Human polynucleotide SEQ ID NO 629.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200190304-A2.
XX
PD 29-NOV-2001.
XX
PF 18-MAY-2001; 2001WO-US16450.
XX
PR 19-MAY-2000; 2000US-205515P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;
XX
DR WPI; 2002-122018/16.
DR P-PSDB; ABB89658.
XX
PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders -
XX
PS Claim 4; SEQ ID NO 629; 2081pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1083 BP; 301 A; 290 C; 199 G; 286 T; 7 other;

Alignment Scores:

Pred. No.:	1.43e-35	Length:	1083
Score:	733.00	Matches:	146
Percent Similarity:	98.65%	Conservative:	0
Best Local Similarity:	98.65%	Mismatches:	2
Query Match:	38.38%	Indels:	1
DB:	24	Gaps:	0

US-09-972-758A-2 (1-359) x ABL90067 (1-1083)

```
Qy        212 AspHisAspGlnGluGluProAspLeuLysThrGlyLeuTyrSerLysArgAlaAlaAla 231
              ||||||||||||||||||||||||||||||||||||||||||||||||||||    |||
Db       1081 GATCACGACCAGGAGGAGCCGGACCTCAAAACCGGCCTKTACTCCAARCGGSCCGCCGC- 1023

Qy        232 LysSerAspAspThrSerAspAspAspPheMetGluGluGlyGlyGluGluAspGlyGly 251
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1022 AAATCCGACGACACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGC 963

Qy        252 SerAspGlyMetGlyGlyAspGlySerGluPheLeuGlnArgAspPheSerGluThrTyr 271
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        962 AGCGATGGGATGGGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTAC 903

Qy        272 GluArgTyrHisThrGluSerLeuGlnAsnMetSerLysGlnGluLeuIleLysGluTyr 291
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        902 GAGCGGTACCACACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTAC 843

Qy        292 LeuGluLeuGluLysCysLeuSerArgMetGluAspGluAsnAsnArgLeuArgLeuGlu 311
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        842 CTGGAACTGGAGAAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGGAG 783

Qy        312 SerLysArgLeuGlyGlyAspAspAlaArgValArgGluLeuGluLeuGluLeuAspArg 331
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        782 AGCAAGCGGCTGGGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGG 723

Qy        332 LeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnGluLeuHisArgGlnGlnGluArg 351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        722 CTGCGCGCCGAGAACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGCGA 663

Qy        352 AlaProLeuSerLysPheGlyAsp 359
              |||||||||||||||||||||||||
Db        662 GCGCCGCTTTCCAAGTTTGGAGAC 639
```

RESULT 6
AAV86036
ID AAV86036 standard; cDNA; 523 BP.
XX
AC AAV86036;
XX
DT 27-APR-1999 (first entry)
XX
DE EST clone B227.
XX
KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
OS Homo sapiens.
XX
PN WO9845435-A2.
XX
PD 15-OCT-1998.
XX
PF 10-APR-1998; 98WO-US06954.
XX
PR 10-APR-1997; 97US-0835913.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX

```
DR   WPI; 1999-070076/06.
XX
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries
XX
PS   Claim 1; Page 97; 633pp; English.
XX
CC   This sequence represents an expressed sequence tag (EST), and is a
CC   polynucleotide of the invention. The polynucleotides of the invention are
CC   all secreted EST sequences isolated from a variety of human tissue
CC   sources. The EST sequences and proteins encoded by them are predicted to
CC   have useful biological activities which would make them suitable for
CC   treating, preventing or ameliorating medical conditions in humans and
CC   animals, although no supporting data is given. Suggested activities
CC   include nutritional activity, immune stimulating or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The EST sequences are also stated to be useful for gene
CC   therapy.
XX
SQ   Sequence 523 BP; 112 A; 149 C; 153 G; 108 T; 1 other;

Alignment Scores:
Pred. No.:              9.34e-35        Length:         523
Score:                  714.00          Matches:        131
Percent Similarity:     99.26%          Conservative:   3
Best Local Similarity:  97.04%          Mismatches:     1
Query Match:            37.38%          Indels:         0
DB:                     20              Gaps:           0

US-09-972-758A-2 (1-359) x AAV86036 (1-523)

Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        104 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 163

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        164 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 223

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        224 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 283

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        284 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 343

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        344 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 403

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        404 GGCGACTTYCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 463

Qy        121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAla 135
              |||||||||||||||||||||||:::||:::   |||:::
Db        464 CCTTGTCATGACTCCGAGGCCAGTAAGATGGGGTCGACGGCCTCT 508

RESULT 7
ABX54966
ID   ABX54966 standard; cDNA; 414 BP.
XX
AC   ABX54966;
XX
DT   26-FEB-2003  (first entry)
XX
DE   Bovine EST associated with lactation/muscle/fat deposition #4895.
XX
KW   Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW   muscle deposition; fat deposition;  genome mapping; gene identification;
KW   gene analysis; cattle breeding.
XX
OS   Bos Taurus.
XX
PN   US2002137160-A1.
XX
PD   26-SEP-2002.
XX
PF   26-OCT-2001; 2001US-0983965.
XX
PR   17-DEC-1998;   98US-113678P.
PR   15-DEC-1999;   99US-0465231.
XX
PA   (BYAT/) BYATT J C.
PA   (MATH/) MATHIALAGAN N.
PA   (TAON/) TAO N.
PA   (WARR/) WARREN W C.
XX
PI   Byatt JC,  Mathialagan N,  Tao N,  Warren WC;
XX
DR   WPI; 2003-102386/09.
XX
PT   Purified nucleic acid molecules, useful for genome mapping, gene
PT   identification and analysis, cattle breeding or preparation of
PT   constructs for cattle gene expression and genetically improved cattle -
XX
PS   Claim 2; SEQ ID No 4895; 38pp; English.
XX
CC   The invention relates to a purified nucleic acid molecule associated with
CC   lactation or muscle and fat deposition (designated LMFD), derived
CC   from cattle, and the LMFD nucleic acid can specifically hybridise to a
CC   second nucleic acid molecule comprising any of 5912 nucleotide
CC   sequences, appearing as ABX50072-ABX55983, or complements of them.
CC   Also included are; (1) a transformed cell having a nucleic acid
CC   comprising an LMFD nucleic acid linked to a promoter and a 3' non-
CC   translated sequence that functions in the cell to cause termination of
CC   transcription and addition of polyadenylated ribonucleotides to a 3' end
CC   of the mRNA molecule; and (2) determining a level or pattern of a
CC   molecule in a bovine cell or tissue comprising: (a) incubating a marker
CC   nucleic acid (comprising any of the 5912 nucleic acid sequences or its
CC   complement or fragment) with a complementary nucleic acid molecule
CC   obtained from the bovine cell or tissue, where hybridisation between the
CC   marker nucleic acid and the complementary nucleic acid permits the
CC   detection of the molecule; and (b) detecting the level or pattern of the
CC   complementary nucleic acid, where the detection of the complementary
CC   nucleic acid is predictive of the level or pattern of the molecule.
CC   The LMFD nucleic acid is used for determining a level or pattern
CC   of a molecule in a bovine cell or tissue. It is useful for genome
CC   mapping, gene identification and analysis, cattle breeding, preparation
CC   of constructs for use in cattle gene expression, or for genetically
CC   improving cattle. The present sequence is one of the 5912 bovine
CC   LMFD EST (expressed sequence tag) nucleic acids.
CC   Note: The present sequence was not shown in the specification but
CC   was obtained in electronic format from the USPTO web site:
CC   seqdata.uspto.gov/sequence.html?DocID=20020137160.
XX
SQ   Sequence 414 BP; 117 A; 106 C; 130 G; 61 T; 0 other;

Alignment Scores:
Pred. No.:              1.76e-33        Length:         414
Score:                  691.00          Matches:        130
Percent Similarity:     95.62%          Conservative:   1
Best Local Similarity:  94.89%          Mismatches:     6
Query Match:            36.18%          Indels:         0
DB:                     25              Gaps:           0

US-09-972-758A-2 (1-359) x ABX54966 (1-414)

Qy        155 ArgArgProSerLysLysLysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGlu 174
              ||||||||||||||||||||||   |||||||||||||   |||||||||||||||
Db          2 AGACGCCCCTCCAAGAAGAAGCGGCTTTGGAAACCGTACTATACGCTGACCTGGGAGGAG 61
```

```
Qy        175 LysLysLysPheAspGluLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPhe 194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 AAGAAAAAGTTCGATGAAAAACAGAGCCTGCGAGCTTCGAGGATTCGAGCCGAGATGTTC 121

Qy        195 AlaLysGlyGlnProValAlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAsp 214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 GCCAAGGGCCAGCCAGTTGCTCCCTATAACACCACGCAGTTCCTCATGGATGACCACGAC 181

Qy        215 GlnGluGluProAspLeuLysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAsp 234
              ||||||||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db        182 CAGGAGGAGCCGGATCTTAAAACCGGCCTCTATCCCAAACGGGCCGCTGCCAAATCCGAC 241

Qy        235 AspThrSerAspAspAspPheMetGluGluGlyGlyGluGluAspGlyGlySerAspGly 254
              ||||||||||||:::||||||||||||||||   ||||||||||||||||||||||||||
Db        242 GACACCAGCGATGAGGACTTTATGGAAGAAGCGGGCGAGGAGGATGGGGGCAGCGACGGG 301

Qy        255 MetGlyGlyAspGlySerGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyr 274
              ||||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db        302 ATGGGAGGAGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGAACTATGAGCGGTAC 361

Qy        275 HisThrGluSerLeuGlnAsnMetSerLysGlnGluLeuIleLysGluTyr 291
              |||   ||||||||||||||||||||||||||||||||||||||||||||
Db        362 CACGCGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAAGAGTAC 412
```

RESULT 8
ABL63543
ID ABL63543 standard; DNA; 461 BP.
XX
AC ABL63543;
XX
DT 15-MAY-2002 (first entry)
XX
DE Breast cancer related gene sequence SEQ ID NO:1880.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
DR WPI; 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
PS Claim 1; SEQ ID 1880; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ Sequence 461 BP; 105 A; 134 C; 138 G; 80 T; 4 other;

Alignment Scores:

Pred. No.:	1.02e-29	Length:	461
Score:	629.00	Matches:	120
Percent Similarity:	93.85%	Conservative:	2
Best Local Similarity:	92.31%	Mismatches:	8
Query Match:	32.93%	Indels:	1
DB:	24	Gaps:	0

US-09-972-758A-2 (1-359) x ABL63543 (1-461)

```
Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         72 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 131

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        132 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGNCACGGAGCGGGTG 191

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 251
```

```
Qy          61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
               |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         252 CCGGACGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 311

Qy          81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         312 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 371

Qy         101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
               ||||||||         :::||||||||||||||||||||||||||||||   |||
Db         372 GC-GACTTCCCGGCNCGGCAGAAAGTGGAGCCGACGCCCGAGGCCGAGCTGCTTGNCCAG 430

Qy         121 ProCysHisAspSerGluAlaSerLysLeu 130
               ||||||||||||||||   |||||||||
Db         431 CCTTGTCATGACTCCGAGGGCAGTAAGTTG 460

RESULT 9
ABQ61099
ID   ABQ61099 standard; cDNA; 1557 BP.
XX
AC   ABQ61099;
XX
DT   26-FEB-2003  (first entry)
XX
DE   Mouse putative protein #15 encoding sequence.
XX
KW   Neuroprotective; immunomodulator; cancer;
KW   cytostatic; anti-inflammatory; gene therapy; nutritional supplement;
KW   wound; burn; ulcer; Alzheimer's disease; Huntington's disease;
KW   amyotrophic lateral sclerosis; autoimmune disorder; inflammation;
KW   vulnerary; gene; ss.
XX
OS   Mus musculus.
XX
PN   WO200231111-A2.
XX
PD   18-APR-2002.
XX
PF   11-OCT-2001; 2001WO-US27760.
XX
PR   12-OCT-2000; 2000US-0687527.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Zhou P,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Wehrman T,  Drmanac RT;
XX
DR   WPI; 2002-426278/45.
DR   N-PSDB; ABP43855.
XX
PT   New polypeptides and their encoded proteins, useful as nutritional
PT   sources or supplements, or in gene therapy, particularly for treating
PT   wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
PT   inflammation  -
XX
PS   Claim 1; SEQ ID # 312; 357pp + sequence listing; English.
XX
CC   The invention relates to 446 newly isolated polynucleotide sequences.
CC   The activity of polynucleotides of the invention may be described as,
CC   vulnerary, neuroprotective, immunomodulator, cytostatic and
CC   anti-inflammatory. Compositions comprising nucleic acids of the invention
CC   are useful for treating a mammalian subject, or as nutritional sources or
CC   supplements. These are useful in gene therapy, particularly for treating
CC   wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
CC   amyotrophic lateral sclerosis, autoimmune disorders, cancer or
CC   inflammation. The nucleic acids and polypeptides are also useful in
CC   diagnostic and research methods. The sequences given in records
CC   ABQ60788-ABQ61233 represent polynucleotides of the invention.
CC   NOTE: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1557 BP; 355 A; 426 C; 513 G; 263 T; 0 other;
```

```
Alignment Scores:
Pred. No.:                3.6e-24        Length:          1557
Score:                    545.50         Matches:         131
Percent Similarity:       57.28%         Conservative:    42
Best Local Similarity:    43.38%         Mismatches:      99
Query Match:              28.56%         Indels:          32
DB:                       24             Gaps:            6

US-09-972-758A-2 (1-359) x ABQ61099 (1-1557)

Qy          58 ArgProGlyProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGln 77
               |||   |||   |||  :::|||   |||  ::::::||||||  :::|||:::
Db         145 AGATTTGGAACAGAAGATGATGGCCA-CTCCGAACCAGACCGCCTGTAATGCAGAGTCAC 203

Qy          78 AlaCysProGluSerSerCysLeuArgGluGlyGluLysGlyGln---------AsnGly 94
                  |||                  |||       |||
Db         204 CAGTGGCCCTG-GAGGAGGCCAAGACCTCTGGTGCCCCGGGGAGCCCCCAAACACCCCCT 262

Qy          95 AspAspSerSerAlaGlyGlyAspPheProProProAlaGluValGluProThrProGlu 114
               :::        :::||||||      |||           :::|||         |||
Db         263 GAGCGTCATGACTCTGGTGGTTCCCTGCCCCTGACACCGCGGATGGAGAGCCACTCAGAG 322

Qy         115 AlaGluLeuLeuAlaGlnProCys-----------HisAspSerGluAlaSerLysLeu 130
                  |||   ||||||                  :::|||   :::   :::
Db         323 GATGAAGATCTTGCTGGGGCTGTCGGTGGCCTGGGCTGGAACAGTAGGAGTCCCCGGACC 382

Qy         131 GlyAlaProAlaAlaGlyGlyGluGluTrpGlyGlnGlnGlnArgGlnLeuGlyLys 150
               :::|||                 |||                        |||   :::
Db         383 CAGAGCCCAGGGGGCTGCTCAGCGGAG-----------------GCTGTGCTGGCCCGG 424

Qy         151 LysLysHisArgArgArgProSerLysLysLysArgHisTrpLysProTyrTyrLysLeu 170
               |||||||||||||||||||||||||:::|||||||||||:::||||||   :::|||
Db         425 AAGAAACACCGTCGGCGGCCATCGAAGCGCAAAAGGCACTGGCGACCCTACCTGGAGCTG 484

Qy         171 ThrTrpGluGluLysLysLysPheAspGluLysGlnSerLeuArgAlaSerArgIleArg 190
               :::|||   ||||||::::::   ||||||:::||||||   |||||||||||:::|||
Db         485 AGCTGGGCTGAGAAACAACAGCGGGATGAGAGGCAGAGCCAGAGGGCCTCCCGGGTCCGC 544

Qy         191 AlaGluMetPheAlaLysGlyGlnProValAlaProTyrAsnThrThrGlnPheLeuMet 210
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         545 GAAGAGATGTTCGCCAAAGGCCAGCCCGTGGCCCCCTACAACACCACCCAGTTCCTGATG 604

Qy         211 AspAspHisAspGlnGluGluPro------AspLeuLysThrGlyLeuTyrSerLysArg 228
               :::|||   |||   |||||||||      |||:::      |||:::
Db         605 AATGACAGGGACCCGGAGGAGCCCAACTTGGATGTGCCCCATGGGATCTCCCACCCAGGT 664

Qy         229 AlaAlaAlaLysSerAspAspThrSerAspAspAspPheMetGluGluGlyGlyGluGlu 248
               ::::::   :::|||                                         |||
Db         665 TCCAGTGGGGAGAGT-----------------------------------------GAG 682

Qy         249 AspGlyGlySerAspGlyMetGlyGlyAspGlySerGluPheLeuGlnArgAspPheSer 268
                  |||   |||||||||   |||             ||||||   ::::::|||||||||
Db         683 GCCGGGGACAGTGATGGGCGGGGCCGAGCGCACGGTGAGTTCCAGCGGAAGGACTTCTCT 742

Qy         269 GluThrTyrGluArgTyrHisThrGluSerLeuGlnAsnMetSerLysGlnGluLeuIle 288
               |||||||||||||:::|||||||||||||||||||      |||||||||||||:::
Db         743 GAGACTTACGAACGCTTCCACACCGAGAGCCTGCAGGGCCGCAGCAAGCAGGAGCTGGTG 802

Qy         289 LysGluTyrLeuGluLeuGluLysCysLeuSerArgMetGluAspGluAsnAsnArgLeu 308
               ::::::|||||||||||||||||   ||||||:::   |||:::|||      ||||||
Db         803 CGAGACTACCTGGAGCTGGAGAAGCGGCTGTCGCAGGCGGAGGAGGAGACTAGGAGGCTG 862

Qy         309 ArgLeuGluSerLysArgLeuGlyGlyAspAspAlaArg---ValArgGluLeuGluLeu 327
               :::                  |||             |||   |||   ||||||
Db         863 CAGCAGCTGCAGGCGTGCACCGGCCAGCAGTCCTGCCGCCAGGTGGAGGAGCTGGCTGCC 922

Qy         328 GluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnGluLeuHisArg 347
               |||:::  |||||||||   ||||||   :::|||      |||||:::::::
Db         923 GAGGTCCAGAGGCTCCGGACCGAAAACCAGCGGCTTCGTCAGGAGAACCAGATGTGGAAC 982

Qy         348 GlnGln 349
               ::::::
Db         983 GAGCAG 988
```

```
Db        983 CGAGAG 988

RESULT 10
AAH99667/c
ID   AAH99667 standard; cDNA; 997 BP.
XX
AC   AAH99667;
XX
DT   16-OCT-2001  (first entry)
XX
DE   Human protein encoding cDNA sequence SEQ ID NO:502.
XX
KW   Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW   antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW   antibacterial; endocrine; cardiant; central nervous system; virucide;
KW   anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW   antiaggregant; haemostatic; vulnerary; antiulcer; osteopathic; eczema;
KW   dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW   neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW   immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW   antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW   cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW   genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW   thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW   allergic rhinitis; diabetes; multiple sclerosis; depression;
KW   Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW   neurological disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153455-A2.
XX
PD   26-JUL-2001.
XX
PF   22-DEC-2000; 2000WO-US35017.
XX
PR   23-DEC-1999;   99US-0471275.
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-457603/49.
DR   P-PSDB; AAM25726.
XX
PT   Isolated human polynucleotides encoding polypeptides, useful for the
PT   treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX
PS   Claim 1; Page 566-567; 1217pp; English.
XX
CC   AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC   AAM25963. The proteins can have activities based on the tissues and
CC   cells they are expressed in, such as: antiinflammatory; antirheumatic;
CC   antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
CC   central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC   cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;
CC   antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
CC   antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC   antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC   encoding them can be used in gene therapy, antisense therapy and vaccine
CC   production, The proteins and polynucleotides are useful for screening for
CC   agonists or antagonists of a protein and for the treatment and diagnosis
CC   of disorders associated with the activity of a protein e.g. inflammation,
CC   rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC   neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC   infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC   anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC   osteoporosis, severe combined immunodeficiency, eczema, allergic
CC   rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC   Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC   neurological disorders.
XX
SQ   Sequence 997 BP; 184 A; 323 C; 266 G; 224 T; 0 other;

Alignment Scores:
Pred. No.:                3.68e-10        Length:          997
Score:                    308.50          Matches:         70
Percent Similarity:       60.26%          Conservative:    21
Best Local Similarity:    46.36%          Mismatches:      43
Query Match:              16.15%          Indels:          17
DB:                       22              Gaps:            3

US-09-972-758A-2 (1-359) x AAH99667 (1-997)

Qy        202 ProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluPro------Asp 219
              ||||||||||||||||||||||||||:::|||   |||   |||||||||      |||
Db        980 CCCTACAACACCACCCAGTTCCTGATGAATGACAGGGACCCGGAGGAGCCCAACTTGGAT 921

Qy        220 LeuLysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAsp 239
              :::     |||:::           ::::::   :::|||
Db        920 GTGCCCCATGGGATCTCCCACCCAGGTTCCAGTGGGGAGAGT----------------- 879

Qy        240 AspPheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGly 259
                                      |||   |||   ||||||||   |||
Db        878 ----------------------GAGGCCGGGGACAGTGATGGGCGGGGCCGAGCGCAC 843

Qy        260 SerGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeu 279
              ||||||   ::::::|||||||||||||||||||||||:::|||||||||||||||
Db        842 GGTGAGTTCCAGCGGAAGGACTTCTCTGAGACTTACGAACGCTTCCACACCGAGAGCCTG 783

Qy        280 GlnAsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSer 299
              |||       |||||||||||||:::::::|||||||||||||||||   ||||||
Db        782 CAGGGCCGCAGCAAGCAGGAGCTGGTGCGAGACTACCTGGAGCTGGAGAAGCGGCTGTCG 723

Qy        300 ArgMetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAsp 319
              :::   |||:::|||     ||||||:::                |||
Db        722 CAGGCGGAGGAGGAGACTAGGAGGCTGCAGCAGCTGCAGGCGTGCACCGGCCAGCAGTCC 663

Qy        320 AlaArg---ValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGln 338
                |||   |||   ||||||      |||:::   |||||||||   ||||||   :::
Db        662 TGCCGCCAGGTGGAGGAGCTGGCTGCCGAGGTCCAGAGGCTCCGGACCGAAAACCAGCGG 603

Qy        339 LeuLeuThrGluAsnGluLeuHisArgGlnGln 349
              |||       ||||||::::::      ::::::
Db        602 CTTCGTCAGGAGAACCAGATGTGGAACCGAGAG 570

RESULT 11
AAV87316
ID   AAV87316 standard; cDNA; 342 BP.
XX
AC   AAV87316;
XX
DT   27-APR-1999  (first entry)
XX
DE   EST clone BP783.
XX
KW   Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW   tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW   chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW   receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
OS   Homo sapiens.
XX
PN   WO9845435-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;   98WO-US06954.
XX
PR   10-APR-1997;   97US-0835913.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
```

```
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-070076/06.
XX
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries
XX
PS   Claim 1; Page 526; 633pp; English.
XX
CC   This sequence represents an expressed sequence tag (EST), and is a
CC   polynucleotide of the invention. The polynucleotides of the invention are
CC   all secreted EST sequences isolated from a variety of human tissue
CC   sources. The EST sequences and proteins encoded by them are predicted to
CC   have useful biological activities which would make them suitable for
CC   treating, preventing or ameliorating medical conditions in humans and
CC   animals, although no supporting data is given. Suggested activities
CC   include nutritional activity, immune stimulating or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The EST sequences are also stated to be useful for gene
CC   therapy.
XX
SQ   Sequence 342 BP; 90 A; 79 C; 83 G; 89 T; 1 other;

Alignment Scores:
Pred. No.:              7.88e-07        Length:         342
Score:                  245.00          Matches:        46
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            12.83%          Indels:         0
DB:                     20              Gaps:           0

US-09-972-758A-2 (1-359) x AAV87316 (1-342)

Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 255

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 315

Qy         41 ProGluGluAspSerArg 46
              |||||||||||||||||
Db        316 CCCGAGGAGGACAGTAGG 333

RESULT 12
ABL06283
ID   ABL06283 standard; cDNA; 1489 BP.
XX
AC   ABL06283;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 13331.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB62180.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 13331; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1489 BP; 434 A; 363 C; 383 G; 309 T; 0 other;

Alignment Scores:
Pred. No.:              0.000257        Length:         1489
Score:                  214.00          Matches:        73
Percent Similarity:     42.20%          Conservative:   46
Best Local Similarity:  25.89%          Mismatches:     94
Query Match:            11.20%          Indels:         69
DB:                     23              Gaps:           10

US-09-972-758A-2 (1-359) x ABL06283 (1-1489)

Qy         85 LeuArgGluGlyGluLysGlyGlnAsnGlyAsp--------------AspSerSerAla 99
                 :::   |||       |||   ::::::|||                    ||||||
Db         98 ATGGCTGAAGCTGTAAAAAATGAAAGTGGCTCCCAACAACGACCTTTGGATAGCGGCGGA 157

Qy        100 GlyGlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAla 119
              ||||||
Db        158 GGAGGGGGC---------------------------------------------------- 166

Qy        120 GlnProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGlu 139
                                    ||||||   |||     |||   ||||||
Db        167 --------------------GCCAGTGGCGGAGGCGGTGTCGCTGTAGGAGGCGGCTCG 205

Qy        140 GluTrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLys 159
                      :::::::::   :::   |||||||||      :::   :::|||   |||
Db        206 GGAATGCCGAAGAGGAAACATCGGCGGGGAAAAAAATCC---AAGATGCAGCCCAAGAAG 262

Qy        160 LysLysArgHisTrpLysProTyrTyrLysLeuThrTrpGlu-----------GluLys 175
                |||   |||:::   |||   :::||||||
Db        263 ACCAAGAACCATTAT---CCGCAGTGGAAACTCGATATGTCAACGGGTGCAGGAGCAACG 319

Qy        176 LysLysPheAspGluLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAla 195
                  :::   :::::::::|||:::   |||         :::|||:::
Db        320 CTGGAGGGGAATCAACGACAGAATAGCAGAACCAAGCTGGTTCGTTCGCGCTCCTTG--- 376

Qy        196 LysGlyGlnProValAlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGln 215
                       :::   ||||||||||||   :::|||||||||:::::|||
Db        377 ----------CTCGTTCCGTACAACACGAACCGCTTCCTAATGGAGGAACACATGTCG 424

Qy        216 GluGluProAspLeuLysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAsp 235
              |||                                                  ||||||
Db        425 GAG-----------------------------------------CTGCACAAAGATGAC 442

Qy        236 ThrSerAspAspAspPheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMet 255
                  :::   |||:::   |||         |||   :::   :::
Db        443 TCCGACGACAACTGTTTT---------GGCTCCCAGACCGAA----------------- 475
```

```
Qy        256 GlyGlyAspGlySerGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHis 275
Db        476 ------GATCAGGTGCTCTTCCTTTCCAAGGAGTTCTCCGATGTCTACGAGCGAGCGCGG 529

Qy        276 ThrGluSerLeuGlnAsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGlu 295
Db        530 CTCGAACGTTTGGAGACGATGAGCAAGCAGGAGCTCATCCAGGAGTGCATGCAAATCGAA 589

Qy        296 LysCysLeuSerArgMetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeu 315
Db        590 GATCGATATTCGAAGGCCCAGAACATATCCAAGGAGTTT---------GGAGCCAAATTA 640

Qy        316 GlyGlyAspAspAlaArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGlu 335
Db        641 CGAGCGCAGGATGATAAGATTCGCCAACTATCACGTGAGAATCAATTTCTGCGCACCCAC 700

Qy        336 AsnLeuGlnLeuLeuThrGluAsnGlu-LeuHisArgGlnGlnGluArgAlaProLeuSe 355
Db        701 CTCTTGCGCACGTGTTCCGCGACAGCTGGACCAGCGGCAGCAGCAGCGGCCGCAGCTTCC 760

Qy        355 rLys 356
Db        761 CCGC 764
```

RESULT 13
ABL06282/c
ID ABL06282 standard; cDNA; 3714 BP.
XX
AC ABL06282;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13328.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB62179.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 13328; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 3714 BP; 971 A; 891 C; 852 G; 1000 T; 0 other;

Alignment Scores:

Pred. No.:	0.00212	Length:	3714
Score:	205.50	Matches:	69
Percent Similarity:	40.36%	Conservative:	42
Best Local Similarity:	25.09%	Mismatches:	85
Query Match:	10.76%	Indels:	79
DB:	23	Gaps:	10

US-09-972-758A-2 (1-359) x ABL06282 (1-3714)

```
Qy         79 CysProGluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAsp--------- 95
Db       2507 TGCAGTGGGCAGAGTTGCATTTTCGAGTATTTTGGCGGTTCAAATTTGAATTTGGTCCTC 2448

Qy         96 ------------------------------------AspSerSerAlaGlyGlyAsp 102
Db       2447 TTGCTCTTTCCAGAAAGTGGCTCCCAACAACGACCTTTGGATAGCGGCGGAGGAGGGGGC 2388

Qy        103 ProProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGlnProCys 122
Db       2388 ------------------------------------------------------------ 2388

Qy        123 HisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGluTrpGly 142
Db       2387 -----------GCCAGTGGCGGAGGCGGTGTCGCTGTAGGAGGCGGCTCGGGAATGCCG 2340

Qy        143 GlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLysLysArg 162
Db       2339 AAGAGGAAACATCGGCGGGGAAAAAAATCC---AAGATGCAGCCCAAGAAGACCAAGAAC 2283

Qy        163 HisTrpLysProTyrTyrLysLeuThrTrpGlu-----------GluLysLysLysPhe 178
Db       2282 CATTAT---CCGCAGTGGAAACTCGATATGTCAACGGGTGCAGGAGCAACGCTGGAGGGG 2226

Qy        179 AspGluLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGln 198
Db       2225 AATCAACGACAGAATAGCAGAACCAAGCTGGTTCGTTCGCGCTCCTTG------------ 2178

Qy        199 ProValAlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluPro 218
Db       2177 ---CTCGTTCCGTACAACACGAACCGCTTCCTAATGGAGGAACACATGTCGGAG------ 2127

Qy        219 AspLeuLysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAsp 238
Db       2126 ----------------------------------CTGCACAAAGATGACTCCGACGAC 2103

Qy        239 AspAspPheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAsp 258
Db       2102 AACTGTTTT---------GGCTCCCAGACCGAA-----------------------GAT 2076

Qy        259 GlySerGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSer 278
Db       2075 CAGGTGCTCTTCCTTTCCAAGGAGTTCTCCGATGTCTACGAGCGAGCGCGGCTCGAACGT 2016

Qy        279 LeuGlnAsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeu 298
Db       2015 TTGGAGACGATGAGCAAGCAGGAGCTCATCCAGGAGTGCATGCAAATCGAAGATCGATAT 1956

Qy        299 SerArgMetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAsp 318
Db       1955 TCGAAGGCCCAGAACATATCCAAGGAGTTT---------GGAGCCAAATTACGAGCGCAG 1905

Qy        319 AspAlaArgValArgGluLeuGluLeuGluLeuAspArgLeuArg 333
Db       1904 GATGATAAGATTCGCCAACTATCACGTGAGAATCAATGTATGCGA 1860
```

RESULT 14
AAC95217
ID AAC95217 standard; cDNA; 549 BP.
XX
AC AAC95217;

XX
DT 19-FEB-2001 (first entry)
XX
DE Cat flea head and nerve cord (HNC) cDNA, SEQ ID NO:1712.
XX
KW Cat flea; head and nerve cord nucleic acid; HNC; flea infestation;
KW vaccine; antiparasitic; therapeutic target; diagnosis; detection; ss.
XX
OS Ctenocephalides felis.
XX
PN WO200061621-A2.
XX
PD 19-OCT-2000.
XX
PF 07-APR-2000; 2000WO-US09437.
XX
PR 09-APR-1999; 99US-0128704.
XX
PA (HESK-) HESKA CORP.
XX
PI Brandt KS, Gaines PJ, Stinchcomb DT, Wisnewski N;
XX
DR WPI; 2000-656323/63.
XX
PT Flea Malpighian tubule and head and nerve cord tissue derived nucleic
PT acids useful for the prevention, diagnosis and treatment of flea
PT infestations -
XX
PS Claim 26; Page 818; 964pp; English.
XX
CC The invention relates to novel cat flea (Ctenocephalides felis) nucleic
CC acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
CC or head and nerve cord (HNC) tissue. The invention also relates to the
CC encoded proteins. The invention additionally encompasses expression
CC constructs, recombinant viruses and recombinant cells comprising the
CC nucleic acids of the invention, recombinant production of the proteins,
CC antibodies against the proteins, a method of identifying inhibitors of
CC the proteins, and compositions comprising the inhibitors for
CC administration to an animal. The nucleic acids, and the proteins they
CC encode may be used in the prevention, treatment and diagnosis of diseases
CC associated with flea infestations. For example, the nucleic acids may be
CC used to produce an HMT or HNC protein according to standard recombinant
CC DNA methodology by inserting the nucleic acids into a host cell and
CC culturing the cell to express the protein. The HMT and HNC nucleic acids
CC may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
CC and quantitate the presence of cat flea or other homologous nucleic acid
CC sequences in samples. They may also be used to study the expression and
CC function of the proteins and their role in metabolism. The HMT and HNC
CC proteins may be used as antigens in the production of specific
CC antibodies, and in assays to identify modulators (agonists and
CC antagonists) of HMT and/or HNC protein expression and activity. The
CC anti-HMT/HNC protein antibodies and antagonists may also be used to
CC downregulate protein expression and activity. The antibodies may also be
CC used as diagnostic agents for detecting the presence of flea polypeptides
CC in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
CC present sequence represents a cat flea HNC cDNA of the invention.
XX
SQ Sequence 549 BP; 207 A; 100 C; 113 G; 126 T; 3 other;

Alignment Scores:

Pred. No.:	0.0058	Length:	549
Score:	184.00	Matches:	66
Percent Similarity:	42.29%	Conservative:	30
Best Local Similarity:	29.07%	Mismatches:	53
Query Match:	9.63%	Indels:	78
DB:	21	Gaps:	12

US-09-972-758A-2 (1-359) x AAC95217 (1-549)

```
Qy         34 ProGlyAlaGluGluArgValProGluGluAspSerArgTrpGlnSerArgAlaPhe--- 52
              |||      ||||||   |||           |||::::::||||||:::   |||
Db         33 CCAGCNGGTGAAGAAATAGTG-----------AGTAAATATCAATCAAAGTGTTTTTTA 80

Qy         53 ---ProGlnLeuGlyGlyArgProGlyProGluGlyGluGlySerLeuGluSerGlnPro 71
                    |||      :::       :::    |||   ::::::||||||:::
Db         81 TTAATACTGTTGAACATCAAAATGAGTGTAAAAATTGAA---AATATTGAATCGGAAAAT 137

Qy         72 Pro-------------------------------------ProLeuGlnThrGlnAla 78
              |||                                    ||||||||||||   |||
Db        138 CCAAAAATAATACAACCGCAGTACAGCGTCCAAAATGTCTGCCCGCTTCAAACAAGCGCT 197

Qy         79 CysProGluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSer 98
                    :::              ||||||||||   :::   |||
Db        198 ---------ACTGAAGCGCCTCTATCAGGAGAAAAAGGAAATGATAGTGAT--------- 239

Qy         99 AlaGlyGlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeu 118

Db        239 ------------------------------------------------------------ 239

Qy        119 AlaGlnProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGlu 138
                                  ||||||:::::::::|||      |||      |||:::
Db        240 ---------------AGAAAAGAAGCAAACCAAGTGGGT------GCAACAAATGGGAAG 278

Qy        139 GluGluTrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSer 158
                                        ||||||||||||||||||
Db        279 --------------------------------AAGAAAAAACATAGACGAGGTAAATGT 305

Qy        159 LysLysLysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPhe 178
              |||:::|||   :::   |||||||||   |||      |||   ::::::|||
Db        306 AAGAGAAAACCTAAT---AAACCGTACAATAAAACAGCCTGGACCCAACGTAAGAATGTC 362

Qy        179 AspGluLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGln 198
                 :::|||:::   :::|||            ||||||::::::   |||   |||
Db        363 ---CAAAAAGAAATGATGAGAGTTCGTAGTGCCCGTGCTAAAATATTAGCCATGGGGCAC 419

Qy        199 ProValAlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluPro 218
                 :::   |||   ||||||   ||||||||||||:::||||||||
Db        420 ACATTAGTTCCTTGTAACACGAACCAATTTCTCATGGAAGATCATGAT------------ 467

Qy        219 AspLeuLysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThr------ 236
                          :::   :::|||   ::::::|||   ||||||   |||
Db        468 --------------GTCCTCACCAAGGATTCATCTGCAGACTCGGACTCTACTTAGTGT 512

Qy        237 ---SerAspAspAspPheMet 242
                 |||:::|||      |||
Db        513 CGTTCTGAAGACACTCTNATG 533
```

RESULT 15
AAS79493
ID AAS79493 standard; cDNA; 1024 BP.
XX
AC AAS79493;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #15297.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.

```
DR   P-PSDB; ABG15306.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 15297; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1024 BP; 238 A; 275 C; 327 G; 183 T; 1 other;

Alignment Scores:
Pred. No.:                0.0437          Length:          1024
Score:                    174.00          Matches:         74
Percent Similarity:       38.06%          Conservative:    36
Best Local Similarity:    25.61%          Mismatches:      87
Query Match:              9.11%           Indels:          94
DB:                       23              Gaps:            12

US-09-972-758A-2 (1-359) x AAS79493 (1-1024)

Qy         26 GluGluLeuAsnProGluArgProProGlyAlaGluGluArg--------------Val 40
              :::::::::::|||         |||      ::::::|||               :::
Db         58 CAGGACATGAGTCCTGGTGAGGGACCCGAGGAGCAGGACAGAGAAGACGGCCCCTGAATG 117

Qy         41 ProGlu----------------------GluAspSerArgTrpGlnSerArgAlaPhe 52
              |||                        |||:::  ||||||:::     :::
Db        118 CCCATTTGGCCCCTTGCTGGCTGGAGGTGTGAAAACCAGCGGTGGAGGCAGCCTTCGGGG 177

Qy         53 Pro--------------------------------GlnLeuGlyGlyArgProGlyPro 61
              |||                              :::|||     |||  |||
Db        178 CCTGCATTTGAGAATAAGGGTGTCACTAGTTCCAGGCGTCTGCT-GAAAGATTTGGAACA 236

Qy         62 GluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro--- 80
              |||   :::|||   |||   ::::::||||||   :::|||:::         |||
Db        237 GAAGATGATGGTCA-CTCCGAACCAGACCGCCTGTAATGCAGAGTCACCAGTGGCCCTGG 295

Qy         81 ------GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSer 98
                    :::     |||                  :::
Db        296 AGGAGGCCAAGACCTCTGGTGTCCCGGGGGAGGCCCCAAACAACCCCTGAGCGTCATGAC 355

Qy         99 AlaGlyGlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeu 118
              :::||||||      |||           :::|||         |||  |||  |||
Db        356 TCTGGTGGTTCCCTGCCCCTGACACCGCGGATGGAGAGCCACTCAGAGGATGAAGATCTT 415

Qy        119 AlaGlnProCys------------HisAspSerGluAlaSerLysLeuGlyAlaProAla 134
              |||                    :::|||   :::   :::      :::|||
Db        416 GCTGGGGCTGTCGGTGGCCTGGGCTGGAACAGTAGGAGTCCCCGGACCCAGAGCCCAGGG 475

Qy        135 AlaGlyGlyGluGluGluTrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArg 154
                     |||                        |||   :::||||||||||||
Db        476 GGCTGCTCAGCGGAG-----------------GCTGTGCTGGCCCGGAAGAAACACCGT 517

Qy        155 ArgArgProSerLysLysLysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGlu 174
              |||||||||||||||:::|||||||||||:::||||||   :::|||:::|||   |||
Db        518 CGGCGGCCATCGAAGCGCAAAAGGCACTGGCGACCCTACCTGGAGCTGAGCTGGGCTGAG 577

Qy        175 LysLysLysPheAspGluLysGlnSer---LeuArgAlaSer---------------- 187
              |||::::::   ||||||:::||||||   |||:::|||
Db        578 AAACAACAGCGGGATGAGAGGCAGAGCNATTTAAAGGCTGGACTAGACGGGTTCTGGGTG 637

Qy        187 ------------------------------------------------------------ 187

Db        638 CCTTGTGCCTCCTTGTTCTTGACAGCCCCCAACCCTGACCCCTCTCTGCAGGGGGTTGGG 697

Qy        188 ---ArgIleArgAlaGluMetPheAla------------LysGlyGlnProValAlaPro 202
                  |||   |||       |||:::              ||||||||||
Db        698 CCTGAGATCCCAGCTGCAGCCTTCAGCTTGGTCACCAAAGACGGGCAGCCTGTTCGC--- 754

Qy        203 TyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeuLysThr 222
                                          :::|||   |||   |||
Db        755 ----------------------------TACGACATGGAGGTGCCC------------ 772

Qy        223 GlyLeuTyrSerLysArgAlaAlaAlaLysSerAsp-----------------AspThr 236
                             ||||||||||||                        |||
Db        773 ------ATGGATTTCCATGCTGCGGCAAAGGGTCAGTGGAGGTTGCAGAGATGCGACCCA 826

Qy        237 SerAspAspAspPheMetGluGluGly 245
                 |||      :::|||   |||
Db        827 GACCCAGACCCAGAACTGGAGATCGGA 853


Search completed: February 5, 2004, 13:48:08
Job time : 404 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 5, 2004, 09:48:27 ; Search time 405 Seconds
(without alignments)
7198.501 Million cell updates/sec

Title: US-09-972-758A-1
Perfect score: 1080
Sequence: 1 atggccgagccattcttgtc.....tttccaagtttggagactag 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1080	100.0	1080	24	ABN84013	Human oestrogen do
	2	1076.8	99.7	2199	19	AAV82778	Clone bp783_3 isol
	3	1076.8	99.7	2199	24	ABQ92015	Human polynucleoti
	4	1075.2	99.6	2178	22	AAH18098	Human cDNA sequenc
c	5	434.2	40.2	1083	24	ABL90067	Human polynucleoti
	6	395.8	36.6	523	20	AAV86036	EST clone B227. H
	7	362.8	33.6	414	25	ABX54966	Bovine EST associa
	8	355.2	32.9	461	24	ABL63543	Breast cancer rela
	9	174.8	16.2	1557	24	ABQ61099	Mouse putative pro
	10	138	12.8	342	20	AAV87316	EST clone BP783.
c	11	112.4	10.4	997	22	AAH99667	Human protein enco
c	12	82.8	7.7	567	22	AAH12885	Human cDNA clone (
	13	68.4	6.3	799	19	AAV55831	Nucleotide sequenc
	14	68.4	6.3	1926	21	AAA50254	Epstein Barr virus
	15	68.4	6.3	1926	22	AAF82902	EBV tethering prot
	16	68.4	6.3	2580	21	AAA75454	Nucleotide sequenc
	17	68.4	6.3	2580	24	AAI64275	Epstein-Barr virus
c	18	68.4	6.3	5452	20	AAX90923	Anti-sense strand
c	19	68.4	6.3	8705	20	AAZ23778	Vector pShuttle DN
	20	68.4	6.3	9600	19	AAV21683	Vector plasmid pCM
c	21	68.4	6.3	10285	24	ABS71027	pCEP-Xa-Fc constru
c	22	68.4	6.3	10285	24	ABS66453	Plasmid pCep-Xa-Fc
	23	68.4	6.3	10380	20	AAZ22248	Nucleotide sequenc
	24	68.4	6.3	10596	14	AAQ51731	Plasmid pCisEBON f
	25	68.4	6.3	10596	17	AAT40348	Plasmid pCisEBON f
	26	68.4	6.3	10596	20	AAX15650	Nucleotide sequenc
c	27	68.4	6.3	16080	21	AAA59553	DNA clone pCEK C1.
	28	67.8	6.3	4270	24	ABN95930	Gene #2428 used to
	29	67.8	6.3	4270	24	ABL64082	Breast cancer rela
	30	67.8	6.3	5250	22	AAK52371	Human polynucleoti
	31	67.8	6.3	5254	22	AAK53355	Human polynucleoti
	32	65.8	6.1	9551	20	AAZ22301	cDNA encoding a hu
	33	65.2	6.0	1925	20	AAX90924	Epstein Barr Virus
	34	65	6.0	795	19	AAV55830	FLGA insert stabil
	35	63.4	5.9	3489	21	AAA30290	Kaposi's sarcoma-a
	36	63.4	5.9	3489	22	AAF82901	Nucleotide sequenc
	37	63.4	5.9	3489	24	ABA93487	Kaposi's sarcoma-a
c	38	63.4	5.9	32207	20	AAV73805	KSHV LUR DNA (nucl
c	39	63.4	5.9	137507	19	AAV19941	KSHV long unique c
	40	60.8	5.6	2108	24	ABL67774	Oesophagus cancer
	41	60.2	5.6	2744	16	AAQ98470	MiSP1-containing p
	42	60.2	5.6	5120	22	AAC84677	DNA sequence of hu
	43	58.4	5.4	1896	21	AAA28684	JEST open reading
	44	58.4	5.4	2313	20	AAZ07197	Human lung tumour
	45	58.4	5.4	2313	21	AAC76151	Human lung tumour-

ALIGNMENTS

RESULT 1
ABN84013
ID ABN84013 standard; cDNA; 1080 BP.
XX
AC ABN84013;
XX
DT 10-SEP-2002 (first entry)
XX
DE Human oestrogen downregulated gene EDG1 coding sequence.
XX
KW EDG1; oestrogen downregulated gene; tumour suppressor; human;
KW breast cancer; prostate cancer; testicular cancer; ovarian cancer;
KW uterine cancer; colon cancer; chromosome 17q; gene therapy; gene;
KW ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1080
FT /*tag= a
FT /product= "EDG1"
FT /transl_except= "(pos:460..462,aa:Arg)"
FT /transl_except= "(pos:511..513,aa:Thr)"
XX
PN WO200228879-A1.
XX
PD 11-APR-2002.
XX
PF 05-OCT-2001; 2001WO-US31300.
XX
PR 05-OCT-2000; 2000US-238187P.

```
XX
PA   (UYCA-) UNIV CASE WESTERN RESERVE.
XX
PI   Montano M,  Wittman B;
XX
DR   WPI; 2002-519107/55.
DR   P-PSDB; ABB76495.
XX
PT   Polynucleotides encoding Estrogen Down-Regulated Gene 1 proteins,
PT   useful for the prevention, diagnosis and treatment of e.g. breast
PT   cancer, testicular cancer, prostate cancer, uterine cancer, cervical
PT   cancer, ovarian cancer and colon cancer -
XX
PS   Claim 1; Fig 1A-B; 52pp; English.
XX
CC   The present sequence is the coding sequence for human oestrogen
CC   downregulated gene 1 (EDG1), a tumour suppressor gene that is
CC   downregulated by oestrogen in mammary epithelial cells.  The gene
CC   was identified by yeast two-hybrid screenings for oestrogen
CC   receptor-interacting proteins in breast epithelial cells.  It
CC   was localised to chromosome arm 17q.  EDG1 mRNA expression is
CC   prevalent in normal mammary epithelial cells and in other human
CC   hormone-responsive tissues such as the ovary, prostate and testis.
CC   Expression is low in breast cancer epithelial cells.  Oestradiol,
CC   which induces breast cancer cell growth, has an inhibitory effect
CC   on EDG1 mRNA expression in breast cancer cells.  Hexamethylene
CC   bis-acetamide, an inducer of differentiation and apoptosis,
CC   upregulates EDG1 mRNA expression in breast cancer cells.  The
CC   invention provides EDG1 polynucleotides and polypeptides.  In a
CC   claimed method, a test sample from an individual suspected of
CC   having, or known to have breast, testicular, prostate, uterine,
CC   cervical, ovarian or colon cancer is assayed for EDG1 transcript
CC   using a polynucleotide that is complementary to the present
CC   sequence or by RT-PCR using a primer derived from the present
CC   sequence.  A decrease in the level of transcript compared to the
CC   level in a test sample indicates that the test sample contains or
CC   was derived from cancerous cells antibody.  A claimed method for
CC   decreasing the proliferation of breast, prostate, testicular,
CC   ovarian, uterine, cervical or colon cancer cells involves increasing
CC   EDG1 protein activity in the cells, either by contacting the cells
CC   with EDG1 protein or its fragment or functional equivalent, or with
CC   a nucleic acid encoding EDG1 protein, its fragment or functional
CC   equivalent.
XX
SQ   Sequence 1080 BP; 265 A; 296 C; 376 G; 143 T; 0 other;

  Query Match             100.0%;  Score 1080;  DB 24;  Length 1080;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-213;
  Matches 1080;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60

Qy         61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120

Qy        121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180

Qy        181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240

Qy        241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300

Qy        301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360

Qy        361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 420

Qy        421 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAG 480

Qy        481 AAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAG 540

Qy        541 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 600

Qy        601 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 660

Qy        661 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 720

Qy        721 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 780

Qy        781 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 840

Qy        841 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 900

Qy        901 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 960

Qy        961 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1020

Qy       1021 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080

RESULT 2
AAV82778
ID   AAV82778 standard; cDNA; 2199 BP.
XX
AC   AAV82778;
XX
DT   25-FEB-1999  (first entry)
XX
DE   Clone bp783_3 isolated from human foetal kidney cDNA library.
XX
KW   Secreted protein; nutritional activity; immune stimulating; vaccine;
KW   suppressing activity; haematopoiesis regulating activity;
KW   tissue growth activity; activin; inhibin activity; chemotactaxis;
KW   chemokinetic activity; haemostasis; thrombolytic activity; receptor;
KW   ligand; anti-inflammatory; cadherin; tumour invasion suppressor;
KW   tumour inhibition; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO9842739-A2.
XX
PD   01-OCT-1998.
XX
```

PF 20-MAR-1998; 98WO-US05653.
XX
PR 19-MAR-1998; 98US-0044466.
PR 21-MAR-1997; 97US-0822167.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX
DR WPI; 1998-609890/51.
DR P-PSDB; AAW85455.
XX
PT New polynucleotides encoding secreted human proteins - derived from
PT human foetal brain, adult brain, foetal kidney, placenta or adult
PT pineal gland cDNA libraries.
XX
PS Claim 1; Page 66-67; 113pp; English.
XX
CC The present sequence encodes a secreted protein. The polynucleotide and
CC secreted protein are predicted to have biological activities which would
CC make them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is given.
CC Suggested activities include nutritional activity, immune stimulating
CC (e.g. as vaccines) or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC invasion suppressor activity, and tumour inhibition activity (no data is
CC given in the specification to support these activities). The
CC polynucleotide is also stated to be useful for gene therapy.
XX
SQ Sequence 2199 BP; 552 A; 511 C; 674 G; 462 T; 0 other;

Query Match 99.7%; Score 1076.8; DB 19; Length 2199;
Best Local Similarity 99.8%; Pred. No. 5.9e-213;
Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        707 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 766

Qy         61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        767 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 826

Qy        121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        827 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 886

Qy        181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        887 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 946

Qy        241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        947 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 1006

Qy        301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1007 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 1066

Qy        361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1067 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 1126

Qy        421 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAG 480
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       1127 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 1186

Qy        481 AAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAG 540
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1187 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 1246

Qy        541 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1247 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 1306

Qy        601 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1307 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 1366

Qy        661 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1367 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 1426

Qy        721 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1427 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 1486

Qy        781 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 1546

Qy        841 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 1606

Qy        901 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 1666

Qy        961 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1726

Qy       1021 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1786
```

RESULT 3
ABQ92015
ID ABQ92015 standard; cDNA; 2199 BP.
XX
AC ABQ92015;
XX
DT 04-OCT-2002 (first entry)
XX
DE Human polynucleotide SEQ ID NO 12.
XX
KW Human; cytostatic; antirheumatic; antiarthritic; vulnerary; analgesic;
KW antiinflammatory; antibacterial; immunosuppressive; antiparkinsonian;
KW neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;
KW antiulcer; fungicide; antidiabetic; antiasthmatic; antiallergic;
KW immunostimulant; antiparasitic; secreted protein; transmembrane protein;
KW cytokine; cell proliferation; cell differentiation; autoimmune disease;
KW stem cell; growth factor; nervous system disease; neuropathy;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW osteoporosis; severe combined immunodeficiency; SCID; infection;
KW multiple sclerosis; rheumatoid arthritis; gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN US2002065394-A1.
XX
PD 30-MAY-2002.
XX
PF 22-DEC-2000; 2000US-0745763.
XX
PR 18-MAR-1998; 98US-0040963.
XX
PA (JACO/) JACOBS K.
PA (MCCO/) MCCOY J M.
PA (LAVA/) LAVALLIE E R.
PA (COLL/) COLLINS-RACIE L A.

PA (EVAN/) EVANS C.
PA (MERB/) MERBERG D.
PA (TREA/) TREACY M.
PA (SPAU/) SPAULDING V.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Spaulding V;
XX
DR WPI; 2002-582343/62.
DR P-PSDB; ABP61799.
XX
PT Novel secreted or transmembrane protein and polynucleotide encoding the
PT protein, useful for diagnosis and treatment of neurological disorders,
PT cancer, autoimmune diseases, bone disorders and lung or liver fibrosis
PT -
XX
PS Claim 37; Page 111-112; 284pp; English.
XX
CC The invention relates to human secreted or transmembrane protein (I),
CC their fragments and is encoded by specific complementary deoxyribonucleic
CC acid (cDNA) inserts (II), where the protein is substantially free from
CC other mammalian proteins. (I) are useful for preventing, treating or
CC ameliorating a medical condition, especially immunological treatment or
CC prevention of tumours. (I) exhibits activity relating to angiogenesis,
CC cytokine, cell proliferation, cell differentiation, antiinflammatory,
CC stem cell growth factor activity and activin or inhibin-related
CC activities. (I) can be used to manipulate stem cells in culture to give
CC rise to neuroepithelial cells that can be used to augment or replace
CC cells damaged by illness, autoimmune disease, accidental damage or
CC genetic disorders. (I) induces the proliferation of neural cells and
CC regeneration of nerve and brain tissue and is useful for the treatment of
CC central and peripheral nervous system diseases and neuropathies, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis. (I) is involved in chemotactic or chemokinetic
CC activity, regulation of haematopoiesis and is useful for treating myeloid
CC or lymphoid cell disorders, platelet disorders such as thrombocytopaenia
CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve
CC tissue growth and in tissue repair, healing of burns, incisions, ulcers,
CC for treating osteoporosis, osteoarthritis, bone degenerative disorders or
CC periodontal disease. (I) is also useful for gut protection or
CC regeneration and treatment of lung or liver fibrosis, reperfusion injury
CC in various tissues, various immune deficiencies and disorders including
CC severe combined immunodeficiency (SCID), bacterial or fungal infections,
CC autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis,
CC diabetes mellitus, myasthenia gravis, allergic reactions and conditions,
CC such as asthma or other respiratory problems. (II) is useful to express
CC recombinant protein, as markers for tissues in which the corresponding
CC protein is preferentially expressed and in gene therapy. The present
CC sequence is that of a polynucleotide of the invention.
XX
SQ Sequence 2199 BP; 552 A; 511 C; 674 G; 462 T; 0 other;

Query Match 99.7%; Score 1076.8; DB 24; Length 2199;
Best Local Similarity 99.8%; Pred. No. 5.9e-213;
Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy        1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      707 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 766

Qy       61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      767 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 826

Qy      121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      827 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 886

Qy      181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      887 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 946

Qy      241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      947 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 1006

Qy      301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1007 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 1066

Qy      361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1067 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 1126

Qy      421 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAG 480
            |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     1127 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 1186

Qy      481 AAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAG 540
            |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     1187 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 1246

Qy      541 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1247 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 1306

Qy      601 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1307 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 1366

Qy      661 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1367 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 1426

Qy      721 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1427 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 1486

Qy      781 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1487 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 1546

Qy      841 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1547 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 1606

Qy      901 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1607 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 1666

Qy      961 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1667 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1726

Qy     1021 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1727 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1786
```

RESULT 4
AAH18098
ID AAH18098 standard; cDNA; 2178 BP.
XX
AC AAH18098;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:17952.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.

```
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 17952; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 2178 BP; 525 A; 513 C; 678 G; 462 T; 0 other;

  Query Match           99.6%;  Score 1075.2;  DB 22;  Length 2178;
  Best Local Similarity  99.7%;  Pred. No. 1.3e-212;
  Matches 1077;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        716 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 775

Qy         61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        776 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 835

Qy        121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        836 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 895

Qy        181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        896 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 955

Qy        241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        956 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 1015

Qy        301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1016 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 1075

Qy        361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1076 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 1135

Qy        421 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAG 480
              ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       1136 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 1195

Qy        481 AAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAG 540
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1196 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 1255

Qy        541 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1256 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 1315

Qy        601 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1316 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 1375

Qy        661 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1376 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 1435

Qy        721 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1436 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 1495

Qy        781 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1496 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 1555

Qy        841 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1556 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 1615

Qy        901 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1616 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 1675

Qy        961 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1676 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1735

Qy       1021 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080
              ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1736 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCCTTCCAAGTTTGGAGACTAG 1795

RESULT 5
ABL90067/c
ID   ABL90067 standard; cDNA; 1083 BP.
XX
AC   ABL90067;
XX
DT   24-MAY-2002  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 629.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200190304-A2.
XX
PD   29-NOV-2001.
```

```
XX
PF   18-MAY-2001; 2001WO-US16450.
XX
PR   19-MAY-2000; 2000US-205515P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Birse CE,  Rosen CA;
XX
DR   WPI; 2002-122018/16.
DR   P-PSDB; ABB89658.
XX
PT   Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT   prevention of neural, immune system, muscular, reproductive,
PT   gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT   disorders -
XX
PS   Claim 4; SEQ ID NO 629; 2081pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC   (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1083 BP; 301 A; 290 C; 199 G; 286 T; 7 other;

  Query Match             40.2%;  Score 434.2;  DB 24;  Length 1083;
  Best Local Similarity   98.9%;  Pred. No. 2.2e-80;
  Matches  444;  Conservative    3;  Mismatches    1;  Indels    1;  Gaps    1;

Qy        632 ATGATCACGACCAGGAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCCG 691
              ||||||||||||||||||||||| ||||||||||||||:||||||||:|||:||||||
Db       1083 ATGATCACGACCAGGAGGAGCCGGACCTCAAAACCGGCCTKTACTCCAARCGGSCCGCCG 1024

Qy        692 CCAAATCCGACGACACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGG 751
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1023 -CAAATCCGACGACACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGG 965

Qy        752 GCAGCGATGGGATGGGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGT 811
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        964 GCAGCGATGGGATGGGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGT 905

Qy        812 ACGAGCGGTACCACACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGT 871
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        904 ACGAGCGGTACCACACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGT 845

Qy        872 ACCTGGAACTGGAGAAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGG 931
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        844 ACCTGGAACTGGAGAAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGG 785

Qy        932 AGAGCAAGCGGCTGGGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACC 991
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        784 AGAGCAAGCGGCTGGGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACC 725

Qy        992 GGCTGCGCGCCGAGAACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGC 1051
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 GGCTGCGCGCCGAGAACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGC 665

Qy       1052 GAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080
              |||||||||||||||||||||||||||||
Db        664 GAGCGCCGCTTTCCAAGTTTGGAGACTAG 636


RESULT 6
AAV86036
ID   AAV86036 standard; cDNA; 523 BP.
XX
AC   AAV86036;
XX
DT   27-APR-1999  (first entry)
XX
DE   EST clone B227.
XX
KW   Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW   tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW   chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW   receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
OS   Homo sapiens.
XX
PN   WO9845435-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;   98WO-US06954.
XX
PR   10-APR-1997;   97US-0835913.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-070076/06.
XX
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries
XX
PS   Claim 1; Page 97; 633pp; English.
XX
CC   This sequence represents an expressed sequence tag (EST), and is a
CC   polynucleotide of the invention. The polynucleotides of the invention are
CC   all secreted EST sequences isolated from a variety of human tissue
CC   sources. The EST sequences and proteins encoded by them are predicted to
CC   have useful biological activities which would make them suitable for
CC   treating, preventing or ameliorating medical conditions in humans and
CC   animals, although no supporting data is given. Suggested activities
CC   include nutritional activity, immune stimulating or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The EST sequences are also stated to be useful for gene
CC   therapy.
XX
SQ   Sequence 523 BP; 112 A; 149 C; 153 G; 108 T; 1 other;

  Query Match             36.6%;  Score 395.8;  DB 20;  Length 523;
  Best Local Similarity   96.6%;  Pred. No. 1.6e-72;
  Matches  403;  Conservative    1;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        104 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 163

Qy         61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        164 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 223

Qy        121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        224 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 283

Qy        181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        284 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 343

Qy        241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        344 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 403

Qy        301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360
              ||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        404 GGCGACTTYCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 463

Qy        361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAG 417
              |||||||||||||||||||||||||| ||||| |  | ||| |   || |   |||
Db        464 CCTTGTCATGACTCCGAGGCCAGTAAGATGGGGTCGACCGCCTCTTTGGCCCTCGAG 520
```

RESULT 7
ABX54966
ID ABX54966 standard; cDNA; 414 BP.
XX
AC ABX54966;
XX
DT 26-FEB-2003 (first entry)
XX
DE Bovine EST associated with lactation/muscle/fat deposition #4895.
XX
KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW muscle deposition; fat deposition; genome mapping; gene identification;
KW gene analysis; cattle breeding.
XX
OS Bos Taurus.
XX
PN US2002137160-A1.
XX
PD 26-SEP-2002.
XX
PF 26-OCT-2001; 2001US-0983965.
XX
PR 17-DEC-1998; 98US-113678P.
PR 15-DEC-1999; 99US-0465231.
XX
PA (BYAT/) BYATT J C.
PA (MATH/) MATHIALAGAN N.
PA (TAON/) TAO N.
PA (WARR/) WARREN W C.
XX
PI Byatt JC, Mathialagan N, Tao N, Warren WC;
XX
DR WPI; 2003-102386/09.
XX
PT Purified nucleic acid molecules, useful for genome mapping, gene
PT identification and analysis, cattle breeding or preparation of
PT constructs for cattle gene expression and genetically improved cattle -
XX
PS Claim 2; SEQ ID No 4895; 38pp; English.
XX
CC The invention relates to a purified nucleic acid molecule associated with
CC lactation or muscle and fat deposition (designated LMFD), derived
CC from cattle, and the LMFD nucleic acid can specifically hybridise to a
CC second nucleic acid molecule comprising any of 5912 nucleotide
CC sequences, appearing as ABX50072-ABX55983, or complements of them.
CC Also included are; (1) a transformed cell having a nucleic acid
CC comprising an LMFD nucleic acid linked to a promoter and a 3' non-
CC translated sequence that functions in the cell to cause termination of
CC transcription and addition of polyadenylated ribonucleotides to a 3' end
CC of the mRNA molecule; and (2) determining a level or pattern of a
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker
CC nucleic acid (comprising any of the 5912 nucleic acid sequences or its
CC complement or fragment) with a complementary nucleic acid molecule
CC obtained from the bovine cell or tissue, where hybridisation between the
CC marker nucleic acid and the complementary nucleic acid permits the
CC detection of the molecule; and (b) detecting the level or pattern of the
CC complementary nucleic acid, where the detection of the complementary
CC nucleic acid is predictive of the level or pattern of the molecule.
CC The LMFD nucleic acid is used for determining a level or pattern
CC of a molecule in a bovine cell or tissue. It is useful for genome
CC mapping, gene identification and analysis, cattle breeding, preparation
CC of constructs for use in cattle gene expression, or for genetically
CC improving cattle. The present sequence is one of the 5912 bovine
CC LMFD EST (expressed sequence tag) nucleic acids.
CC Note: The present sequence was not shown in the specification but
CC was obtained in electronic format from the USPTO web site:
CC seqdata.uspto.gov/sequence.html?DocID=20020137160.
XX
SQ Sequence 414 BP; 117 A; 106 C; 130 G; 61 T; 0 other;

```
  Query Match             33.6%;  Score 362.8;  DB 25;  Length 414;
  Best Local Similarity   92.3%;  Pred. No. 9.9e-66;
  Matches  382;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy        462 GAGACGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGA 521
              ||||||||| |||||||||||||||| |||||||||||||| | ||||| |||||| ||
Db          1 GAGACGCCCCTCCAAGAAGAAGCGGCTTTGGAAACCGTACTATACGCTGACCTGGGAGGA 60

Qy        522 GAAGAAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTT 581
              |||||||||||||| || |||||||||| |||||||| ||||| ||||||||||||||
Db         61 GAAGAAAAAGTTCGATGAAAAACAGAGCCTGCGAGCTTCGAGGATTCGAGCCGAGATGTT 120

Qy        582 CGCCAAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGA 641
              |||||||||||||||| || || ||||||||||||||||||||||||||||||| |||||
Db        121 CGCCAAGGGCCAGCCAGTTGCTCCCTATAACACCACGCAGTTCCTCATGGATGACCACGA 180

Qy        642 CCAGGAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGA 701
              |||||||||||||||| |||||||||| ||  |||| |||||||| ||||||||||
Db        181 CCAGGAGGAGCCGGATCTTAAAACCGGCCTCTATCCCAAACGGGCCGCTGCCAAATCCGA 240

Qy        702 CGACACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGG 761
              |||||||||||||||| ||||| |||||||||| ||| ||||||||||||||||||| ||
Db        241 CGACACCAGCGATGAGGACTTTATGGAAGAAGCGGGCGAGGAGGATGGGGGCAGCGACGG 300

Qy        762 GATGGGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTA 821
              ||||||||| ||||||||||||||||||||||||||||||||  || ||||||||
Db        301 GATGGGAGGAGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGAACTATGAGCGGTA 360

Qy        822 CCACACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTACCT 875
              |||| |||||||||||||||||||||||||||||||||||| ||||||||
Db        361 CCACGCGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAAGAGTACCT 414
```

RESULT 8
ABL63543
ID ABL63543 standard; DNA; 461 BP.
XX
AC ABL63543;
XX
DT 15-MAY-2002 (first entry)
XX
DE Breast cancer related gene sequence SEQ ID NO:1880.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.

PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
DR WPI; 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
PS Claim 1; SEQ ID 1880; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilm's tumour.
XX

SQ Sequence 461 BP; 105 A; 134 C; 138 G; 80 T; 4 other;

Query Match 32.9%; Score 355.2; DB 24; Length 461;
Best Local Similarity 95.7%; Pred. No. 3.7e-64;
Matches 374; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

```
Qy          1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         72 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 131

Qy         61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||    ||||||||||
Db        132 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGNCACGGAGCGGGTG 191

Qy        121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 251

Qy        181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 CCGGACGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 311

Qy        241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        312 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 371

Qy        301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360
              |||||||||||  | | |     |||||||  ||||||||||||||||||||||| | ||||
Db        372 -GCGACTTCCCGGCNCGGCAGAAAGTGGAGCCGACGCCCGAGGCCGAGCTGCTTGNCCAG 430

Qy        361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGG 391
              |||||||||||||||||| ||||||||||
Db        431 CCTTGTCATGACTCCGAGGGCAGTAAGTTGG 461
```

RESULT 9
ABQ61099
ID ABQ61099 standard; cDNA; 1557 BP.
XX
AC ABQ61099;
XX
DT 26-FEB-2003 (first entry)
XX
DE Mouse putative protein #15 encoding sequence.
XX
KW Neuroprotective; immunomodulator; cancer;
KW cytostatic; anti-inflammatory; gene therapy; nutritional supplement;
KW wound; burn; ulcer; Alzheimer's disease; Huntington's disease;
KW amyotrophic lateral sclerosis; autoimmune disorder; inflammation;
KW vulnerary; gene; ss.
XX
OS Mus musculus.
XX
PN WO200231111-A2.
XX
PD 18-APR-2002.
XX
PF 11-OCT-2001; 2001WO-US27760.
XX
PR 12-OCT-2000; 2000US-0687527.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR WPI; 2002-426278/45.
DR N-PSDB; ABP43855.
XX
PT New polypeptides and their encoded proteins, useful as nutritional
PT sources or supplements, or in gene therapy, particularly for treating
PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
PT inflammation -

```
XX
PS   Claim 1; SEQ ID # 312; 357pp + sequence listing; English.
XX
CC   The invention relates to 446 newly isolated polynucleotide sequences.
CC   The activity of polynucleotides of the invention may be described as,
CC   vulnerary, neuroprotective, immunomodulator, cytostatic and
CC   anti-inflammatory. Compositions comprising nucleic acids of the invention
CC   are useful for treating a mammalian subject, or as nutritional sources or
CC   supplements. These are useful in gene therapy, particularly for treating
CC   wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
CC   amyotrophic lateral sclerosis, autoimmune disorders, cancer or
CC   inflammation. The nucleic acids and polypeptides are also useful in
CC   diagnostic and research methods. The sequences given in records
CC   ABQ60788-ABQ61233 represent polynucleotides of the invention.
CC   NOTE: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1557 BP; 355 A; 426 C; 513 G; 263 T; 0 other;

  Query Match             16.2%;  Score 174.8;  DB 24;  Length 1557;
  Best Local Similarity   60.5%;  Pred. No. 7.8e-27;
  Matches  362;  Conservative    0;  Mismatches  197;  Indels   39;  Gaps    3;

Qy        441 GCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTA 500
              |||||     ||| |||||      | || || || |||    ||  |||| ||| |||||
Db        415 GCTGGCCCGGAAGAAACACCGTCGGCGGCCATCGAAGCGCAAAAGGCACTGGCGACCCTA 474

Qy        501 CTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTC 560
              |     |||||| |||||  |||||  || ||    || ||||  |||||||    | || ||
Db        475 CCTGGAGCTGAGCTGGGCTGAGAAACAACAGCGGGATGAGAGGCAGAGCCAGAGGGCCTC 534

Qy        561 AAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTCGCGCCCTATAACACCACGCA 620
                || |||| |  |||||||||||||| |||||||| || || ||||| |||||||| ||
Db        535 CCGGGTCCGCGAAGAGATGTTCGCCAAAGGCCAGCCCGTGGCCCCCTACAACACCACCCA 594

Qy        621 GTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAA 680
              |||||| ||| ||||    |||| |||||||||          ||| |  ||| | |||
Db        595 GTTCCTGATGAATGACAGGGACCCGGAGGAGCC--------CAACTTGGATGTGCCCCAT 646

Qy        681 GCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGACTTCATGGAAGAAGGGGGTGA 740
              | |  |  || || |        |                              ||||
Db        647 GGGATCTCCCACCCAGGTTCCAG----------------------------TGGGGAGAG 678

Qy        741 GGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTT 800
               ||||  |||| ||| |||||| ||| | |    | | ||||| | || |  ||||||
Db        679 TGAGGCCGGGGACAGTGATGGGCGGGGCCGAGCGCACGGTGAGTTCCAGCGGAAGGACTT 738

Qy        801 CTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCT 860
              ||| ||||| ||||| || | |||||| ||||||||||    |   ||||||||||||||
Db        739 CTCTGAGACTTACGAACGCTTCCACACCGAGAGCCTGCAGGGCCGCAGCAAGCAGGAGCT 798

Qy        861 CATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGCATGGAGGACGAGAACAACCG 920
                |    || |||||||| ||||||||| | || ||||    |||||| ||||  |   |
Db        799 GGTGCGAGACTACCTGGAGCTGGAGAAGCGGCTGTCGCAGGCGGAGGAGGAGACTAGGAG 858

Qy        921 GCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGA---CGACGCGCGTGTGCGGGAGCTGGA 977
              ||||| || |  |     | |    ||      |   |  |   |  |||  ||||||||
Db        859 GCTGCAGCAGCTGCAGGCGTGCACCGGCCAGCAGTCCTGCCGCCAGGTGGAGGAGCTGGC 918

Qy        978 GCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTGACCGAGAACGAACTG 1035
                  ||| |  |  |||| ||  |||| ||||  | ||| |     ||||||  |  ||
Db        919 TGCCGAGGTCCAGAGGCTCCGGACCGAAAACCAGCGGCTTCGTCAGGAGAACCAGATG 976


RESULT 10
AAV87316
ID   AAV87316 standard; cDNA; 342 BP.
XX
AC   AAV87316;
XX
DT   27-APR-1999  (first entry)
XX
DE   EST clone BP783.
XX
KW   Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW   tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW   chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW   receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
OS   Homo sapiens.
XX
PN   WO9845435-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;   98WO-US06954.
XX
PR   10-APR-1997;   97US-0835913.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-070076/06.
XX
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries
XX
PS   Claim 1; Page 526; 633pp; English.
XX
CC   This sequence represents an expressed sequence tag (EST), and is a
CC   polynucleotide of the invention. The polynucleotides of the invention are
CC   all secreted EST sequences isolated from a variety of human tissue
CC   sources. The EST sequences and proteins encoded by them are predicted to
CC   have useful biological activities which would make them suitable for
CC   treating, preventing or ameliorating medical conditions in humans and
CC   animals, although no supporting data is given. Suggested activities
CC   include nutritional activity, immune stimulating or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The EST sequences are also stated to be useful for gene
CC   therapy.
XX
SQ   Sequence 342 BP; 90 A; 79 C; 83 G; 89 T; 1 other;

  Query Match             12.8%;  Score 138;  DB 20;  Length 342;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-19;
  Matches  138;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 255

Qy         61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 315

Qy        121 CCCGAGGAGGACAGTAGG 138
              ||||||||||||||||||
Db        316 CCCGAGGAGGACAGTAGG 333


RESULT 11
AAH99667/c
ID   AAH99667 standard; cDNA; 997 BP.
XX
AC   AAH99667;
XX
DT   16-OCT-2001  (first entry)
XX
DE   Human protein encoding cDNA sequence SEQ ID NO:502.
```

```
XX
KW   Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW   antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW   antibacterial; endocrine; cardiant; central nervous system; virucide;
KW   anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW   antiaggregant; haemostatic; vulnerary; antiulcer; osteopathic; eczema;
KW   dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW   neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW   immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW   antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW   cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW   genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW   thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW   allergic rhinitis; diabetes; multiple sclerosis; depression;
KW   Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW   neurological disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153455-A2.
XX
PD   26-JUL-2001.
XX
PF   22-DEC-2000; 2000WO-US35017.
XX
PR   23-DEC-1999;   99US-0471275.
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-457603/49.
DR   P-PSDB; AAM25726.
XX
PT   Isolated human polynucleotides encoding polypeptides, useful for the
PT   treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX
PS   Claim 1; Page 566-567; 1217pp; English.
XX
CC   AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC   AAM25963. The proteins can have activities based on the tissues and
CC   cells they are expressed in, such as: antiinflammatory; antirheumatic;
CC   antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
CC   central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC   cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;
CC   antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
CC   antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC   antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC   encoding them can be used in gene therapy, antisense therapy and vaccine
CC   production, The proteins and polynucleotides are useful for screening for
CC   agonists or antagonists of a protein and for the treatment and diagnosis
CC   of disorders associated with the activity of a protein e.g. inflammation,
CC   rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC   neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC   infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC   anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC   osteoporosis, severe combined immunodeficiency, eczema, allergic
CC   rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC   Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC   neurological disorders.
XX
SQ   Sequence 997 BP; 184 A; 323 C; 266 G; 224 T; 0 other;

  Query Match             10.4%;  Score 112.4;  DB 22;  Length 997;
  Best Local Similarity   62.7%;  Pred. No. 5.4e-14;
  Matches  192;  Conservative    0;  Mismatches  111;  Indels    3;  Gaps    1;

Qy        733 GGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGCGAGTTTCTGCAG 792
              ||||     ||||  |||| ||| ||||||  |||  | |    | | ||||| | || |
Db        887 GGGGAGAGTGAGGCCGGGGACAGTGATGGGCGGGGCCGAGCGCACGGTGAGTTCCAGCGG 828

Qy        793 CGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAGAACATGAGCAAG 852
              ||||||||| ||||| ||||| || | |||||| |||||||||||  |   ||||||
Db        827 AAGGACTTCTCTGAGACTTACGAACGCTTCCACACCGAGAGCCTGCAGGGCCGCAGCAAG 768

Qy        853 CAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGCATGGAGGACGAG 912
              |||||||  |    || |||||||| |||||||| | || ||||    |||||| |||
Db        767 CAGGAGCTGGTGCGAGACTACCTGGAGCTGGAGAAGCGGCTGTCGCAGGCGGAGGAGGAG 708

Qy        913 AACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGA---CGACGCGCGTGTGCGG 969
              | |   |||||| || |  |     | |    ||     |   | | |   | ||| |
Db        707 ACTAGGAGGCTGCAGCAGCTGCAGGCGTGCACCGGCCAGCAGTCCTGCCGCCAGGTGGAG 648

Qy        970 GAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTGACCGAGAAC 1029
              |||||||     ||| |  | |||| ||  |||| |||| | ||| |     ||||||
Db        647 GAGCTGGCTGCCGAGGTCCAGAGGCTCCGGACCGAAAACCAGCGGCTTCGTCAGGAGAAC 588

Qy       1030 GAACTG 1035
                | ||
Db        587 CAGATG 582


RESULT 12
AAH12885/c
ID   AAH12885 standard; cDNA; 567 BP.
XX
AC   AAH12885;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA clone (3'-primer) SEQ ID NO:9720.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 3; SEQ ID 9720; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
```

CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 567 BP; 182 A; 139 C; 87 G; 149 T; 10 other;

Query Match 7.7%; Score 82.8; DB 22; Length 567;
Best Local Similarity 76.7%; Pred. No. 6.3e-08;
Matches 122; Conservative 0; Mismatches 34; Indels 3; Gaps 2;

```
Qy        922 CTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTG 981
              | |   | |  |||  |     | ||  | ||  |||  | | |||||||| |||||  ||
Db        539 CNGNTTCGGTTGAGACAANCNGTTGGGTGGGAAAACGGNCTTTTGCGGGAG-TGGAGTTG 481

Qy        982 GAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTGACCGAGAACGAACTGCACCGG 1041
              |||  ||   | | ||||||||||||||| |||||  || ||||||||||||||  ||||||||
Db        480 GAGTTG--ACCGGTGCGCGCCGAGAACTTCCAGTTGNTGACCGAGAACGAATTGCACCGG 423

Qy       1042 CAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080
              |||||||||||||||||| ||||||||||||||||||||
Db        422 CAGCAGGAGCGAGCGCCGCNTTCCAAGTTTGGAGACTAG 384
```

RESULT 13
AAV55831
ID AAV55831 standard; DNA; 799 BP.
XX
AC AAV55831;
XX
DT 18-NOV-1998 (first entry)
XX
DE Nucleotide sequence of the stabilising sequence-encoding insert.
XX
KW Fusion protein; stabilising polypeptide; proteolytic degradation;
KW resistance; half-life; autoimmune disease; inflammation; nitro drug;
KW IkappaB regulator protein; inflammatory bowel disease; in vivo imaging;
KW nitroreductase protein; enzyme therapy; prodrug therapy; protease;
KW cancer; pathological condition; ss.
XX
OS Epstein-barr virus.
XX
PN WO9822577-A1.
XX
PD 28-MAY-1998.
XX
PF 17-NOV-1997; 97WO-IB01508.
XX
PR 25-JUN-1997; 97US-0048945.
PR 15-NOV-1996; 96US-0030986.
XX
PA (MASU/) MASUCCI M G.
XX
PI Masucci MG;
XX
DR WPI; 1998-312463/27.
XX
PT New fusion proteins resistant to proteolytic degradation -
PT comprising a core protein with a stabilising polypeptide comprising
PT a peptide sequence containing glycine repeats
XX
PS Disclosure; Fig 4B; 120pp; English.
XX
CC This is a nucleotide sequence of the stabilising sequence-encoding
CC insert. The invention provides a method for increasing the resistance
CC of a core protein to proteolytic degradation that comprises linking or
CC inserting onto or into the core protein a stabilising polypeptide of
CC formula [(Glya)X(Glyb)Y(Glyc)Z]n where Glya, Glyb, Glyc are 1-6
CC sequential Gly residues and X, Y, Z are Ala, Ser, Val, Ile, Leu, Met,
CC Phe, Pro or Thr and n can be anything between 1-66. X, Y and Z need not
CC be identical from n repeat to n repeat. Alternatively a nucleic acid
CC encoding the stabilising polypeptide can be linked onto or inserted into
CC a nucleic acid encoding a core protein. The fusion proteins of the
CC invention are more resistant to degradation by proteases and, thus, have
CC a longer half-life than the unfused core protein. The products can be
CC used for treating autoimmune diseases, cancer and inflammation. In
CC particular, the core protein may be an IkappaB regulator protein for the
CC treatment of inflammatory bowel disease, or a nitroreductase protein
CC which can activate nitro drugs in enzyme/prodrug therapy to treat cancer
CC or other pathological conditions. The fusion proteins can also be used in
CC diagnostic methods such as in vivo imaging.
XX
SQ Sequence 799 BP; 201 A; 106 C; 479 G; 13 T; 0 other;

Query Match 6.3%; Score 68.4; DB 19; Length 799;
Best Local Similarity 44.0%; Pred. No. 6.3e-05;
Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

```
Qy        406 GGGGGCGAAGAGGAGTGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGA 465
              |||   ||  |||||  ||| ||| |||||    ||  |  | ||    |  |    |
Db         63 GGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 122

Qy        466 CGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAG 525
                |   || || ||  |     | | | | |   || | |  || | |||  |
Db        123 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGG 182

Qy        526 AAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCC 585
                |  ||   || |||   |||   |     |       |     |    |||  |   |
Db        183 CAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAG 242

Qy        586 AAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAG 645
                |    ||  |   | |   |  |     | ||   |    | | ||  | |  |||
Db        243 GGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG 302

Qy        646 GAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGAC 705
              |||| || ||| |   |     || | | |     ||  ||     |   |    || |
Db        303 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGG 362

Qy        706 ACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATG 765
                | ||  |        ||   | | |||  ||||| | | |||    ||   |  |
Db        363 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 422

Qy        766 GGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCAC 825
              |  |||  | |  .|  ||     ||||  | |       | ||  | | |||  ||  |  |
Db        423 GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 482

Qy        826 ACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAG 885
                |||| | | | ||  |   | |  || ||| ||       |||||   | |||   | ||
Db        483 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAG 542

Qy        886 AAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTG 945
                ||        ||   || | | ||| |  |   |  || |  || |    ||| |  |
Db        543 GAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG 602

Qy        946 GGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAG 1005
              |  ||  | || | ||  | || ||||  | ||  || || ||| | |  |    |||
Db        603 GAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 662

Qy       1006 AACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCG 1059
                |  ||| |  |    || | |  | | | | || ||| ||  | || |
Db        663 GGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAG 716
```

RESULT 14
AAA50254
ID AAA50254 standard; DNA; 1926 BP.
XX
AC AAA50254;
XX
DT 07-NOV-2000 (first entry)

XX
DE Epstein Barr virus nuclear antigen 1 protein (EBNA1) DNA.
XX
KW EBV; nuclear antigen 1; EBNA1; episome; transfection; selection;
KW gene therapy; ds.
XX
OS Epstein-barr virus.
XX
PN WO200047778-A1.
XX
PD 17-AUG-2000.
XX
PF 11-FEB-2000; 2000WO-US03547.
XX
PR 11-FEB-1999; 99US-0249585.
XX
PA (PHAR-) PHARMACOPEIA INC.
XX
PI Horlick RA, Chelsky D;
XX
DR WPI; 2000-515062/46.
DR P-PSDB; AAY95856.
XX
PT Stably transfecting eukaryotic cells with at least one episome for the
PT production of a desired protein in vitro and for gene therapy -
XX
PS Disclosure; Fig 2; 53pp; English.
XX
CC The present sequence is that of DNA encoding the Epstein-Barr virus
CC (EBV) nuclear antigen protein 1 (EBNA1, see AAY95856). EBNA1 is
CC utilised in a novel method for obtaining a eukaryotic cell that is
CC stably transfected with at least one episome. This method involves
CC transfecting a eukaryotic cell with: (1) a first episome comprising
CC an EBV origin of replication (oriP, see AAA50253), a gene encoding a
CC first protein whose expression results in cell death and a
CC selectable marker for eukaryotic cells; and (2) a second episome
CC comprising an EBV oriP and a gene encoding a second protein, where
CC expression of the second protein prohibits the occurrence of cell
CC death resulting from expression of the first protein to produce
CC doubly transfected cells which also express an antigen that
CC promotes retention of the episomes by the cells. The doubly
CC transfected cells are maintained under conditions in which the
CC first and second proteins and the selectable marker are expressed,
CC and the selective pressure specified by the marker is maintained.
CC Under these conditions, only cells containing both episomes live.
CC Preferably, EBNA1 is expressed from 1 of the episomes, and the
CC protein of interest from the other episome. Either or both
CC epitopes may further comprise a nucleic acid sequence encoding a
CC protein desired to be expressed in the cell (e.g. a therapeutic
CC protein), a nucleic acid encoding an RNA that is not intended to
CC be translated (e.g. a therapeutic RNA), or a DNA sequence used as
CC a tag for the cells. The method is applicable to cell culture or
CC intact organisms, for gene therapy. It allows the rapid
CC establishment of eukaryotic cells that stably and reliably express
CC a gene of interest, using a novel method of selection, and
CC maintenance of that selection without the need for exogenous
CC selection factors, such as antibiotics.
XX
SQ Sequence 1926 BP; 487 A; 352 C; 872 G; 215 T; 0 other;

Query Match 6.3%; Score 68.4; DB 21; Length 1926;
Best Local Similarity 44.0%; Pred. No. 7.3e-05;
Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

```
Qy        406 GGGGGCGAAGAGGAGTGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGA 465
              |||   ||  |||||  ||| ||| |||||    ||  |  | ||    |  |   |
Db        305 GGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 364

Qy        466 CGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAG 525
                  |    || || ||  |     | | | | |  |   || | |  || | |||  |
Db        365 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGG 424

Qy        526 AAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCC 585
```

```
              |  ||   || |||   |||   |     |       |    |    |||  |   |
Db        425 CAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAG 484

Qy        586 AAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAG 645
                |   ||  |  | |  |  |    | ||   |   | | ||  | |  |||
Db        485 GGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG 544

Qy        646 GAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGAC 705
              |||| || ||| |   |    || | | |     ||  ||      |   |   || |
Db        545 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGG 604

Qy        706 ACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATG 765
                  | || |        ||   | | |||  |||| | | |||    ||   |  |
Db        605 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 664

Qy        766 GGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCAC 825
              |  ||| | |   |  ||     |||  | |      | ||  | | ||| || |  |
Db        665 GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 724

Qy        826 ACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAG 885
                |||| | | | || |   | |  || ||| ||      |||||   | |||   | ||
Db        725 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAG 784

Qy        886 AAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTG 945
                ||          ||  | | ||| |   |   | || |  || |    ||| |  |
Db        785 GAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG 844

Qy        946 GGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAG 1005
              |  ||  | ||  | ||  | ||  |||| | ||  ||  ||  ||| | |  | |    |||
Db        845 GAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 904

Qy       1006 AACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCG 1059
                |   ||| |  |    || | |  | | | | | || ||| ||  | || |
Db        905 GGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAG 958
```

RESULT 15
AAF82902
ID AAF82902 standard; DNA; 1926 BP.
XX
AC AAF82902;
XX
DT 29-JUN-2001 (first entry)
XX
DE EBV tethering protein EBNA1 encoding DNA.
XX
KW Histone H1; tethering protein; LANA; gene therapy; multiple sclerosis;
KW Parkinson's disease; Huntington disease; diabetes; human herpesvirus 8;
KW EBV; latency-associated nuclear antigen; LANA; EBNA1; ds.
XX
OS Epstein-barr virus.
XX
FH Key Location/Qualifiers
FT CDS 1..1926
FT /*tag= a
XX
PN WO200125484-A2.
XX
PD 12-APR-2001.
XX
PF 29-SEP-2000; 2000WO-US26908.
XX
PR 01-OCT-1999; 99US-0410399.
XX
PA (UNMI) UNIV MICHIGAN.
XX
PI Robertson ES, Cotter MA;
XX
DR WPI; 2001-281736/29.
DR P-PSDB; AAB62332.
XX
PT A composition for use in gene therapy comprises an expression vector
PT that includes a nucleic acid sequence encoding a nucleic acid binding

PT protein -
XX
PS Disclosure; Fig 9C; 60pp; English.
XX
CC The invention provides a composition comprising nucleic acid, histone H1
CC protein and expression vector operationally encoding a protein suitable
CC for tethering the nucleic acid to the histone H1 protein, where the
CC tethering protein is LANA. The composition is useful in aiding the
CC retention of the viral DNA in the host cell. The viral vector encodes a
CC protein suitable for tethering DNA to Histone H1. Methods for screening
CC for compounds which are agonistic or antagonistic for the tethering of
CC viral proteins to histone H1 and DNA binding sites are useful for
CC developing the method of viral transfer. The composition has applications
CC to gene therapy, including the treatment of multiple sclerosis,
CC Parkinson's disease, Huntington disease and diabetes. The present
CC sequence represents the nucleotide sequence of the Epstein-barr virus
CC (EBV) tethering protein ENNA1.
XX
SQ Sequence 1926 BP; 487 A; 352 C; 872 G; 215 T; 0 other;

Query Match 6.3%; Score 68.4; DB 22; Length 1926;
Best Local Similarity 44.0%; Pred. No. 7.3e-05;
Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

```
Qy        406 GGGGGCGAAGAGGAGTGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGA 465
              |||   ||  |||||  ||| ||| |||||     |||  | ||     |  |   |
Db        305 GGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 364

Qy        466 CGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAG 525
                  |    || || ||  |      | | | | | |  | || | |  || | |||  |
Db        365 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGG 424

Qy        526 AAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCC 585
               |  ||   || |||   |||   |     |       |    |    |||  |   |
Db        425 CAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAG 484

Qy        586 AAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAG 645
               |    ||  |   | |     |  |       | ||   |    | | ||  | |  |||
Db        485 GGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG 544

Qy        646 GAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGAC 705
              |||| || ||| |    |     || | | |      ||  ||     |   |    || |
Db        545 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGG 604

Qy        706 ACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATG 765
                | || |       ||   | | |||  ||||| | | |||    ||   |  |
Db        605 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 664

Qy        766 GGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCAC 825
              |  ||| | |   |  ||     ||||  | |       | ||  | | ||| || |  |
Db        665 GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 724

Qy        826 ACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAG 885
                |||| | | | || |   | |  || ||| ||      |||||   | |||   | ||
Db        725 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 784

Qy        886 AAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTG 945
               ||        ||    || | | ||| |  |   |  || |  || |    ||| |  |
Db        785 GAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG 844

Qy        946 GGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAG 1005
              |  ||  | || | ||  | || ||||  | ||  || || ||| | |  | |    |||
Db        845 GAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 904

Qy       1006 AACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCG 1059
                |  ||| |  |    || | |  | | | | | || ||| ||  | || |
Db        905 GGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAG 958
```

Search completed: February 5, 2004, 10:08:37
Job time : 408 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 5, 2004, 09:48:27 ; Search time 4185 Seconds
(without alignments)
10557.323 Million cell updates/sec

Title: US-09-972-758A-1
Perfect score: 1080
Sequence: 1 atggccgagccattcttgtc.....tttccaagtttggagactag 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1076.8	99.7	2199	6	BD106410	BD106410 Secreted
2		1076.8	99.7	3624	9	AB021179	AB021179 Homo sapi
3	c	1076.8	99.7	35050	9	AC142472	AC142472 Homo sapi
4		1076.8	99.7	38849	9	AC138150	AC138150 Homo sapi
5	c	1076.8	99.7	100836	2	AC087298	AC087298 Homo sapi
6		1075.2	99.6	2178	6	BD160090	BD160090 Primer fo
7		1075.2	99.6	2178	9	AK023624	AK023624 Homo sapi
8		1073.6	99.4	2086	9	BC006460	BC006460 Homo sapi
9	c	1052.8	97.5	142326	2	AC024047	AC024047 Homo sapi
10		840.8	77.9	3402	10	AY090614	AY090614 Mus muscu
11		840.8	77.9	111002	10	AL731805	AL731805 Mouse DNA
12		833.8	77.2	1488	10	BC022111	BC022111 Mus muscu
13	c	818	75.7	213625	2	AC120950	AC120950 Rattus no
14	c	471.2	43.6	111803	2	AC136172	AC136172 Rattus no
15		395.8	36.6	523	6	BD058159	BD058159 Secreted
16		355.2	32.9	461	6	AX331371	AX331371 Sequence
17	c	176.8	16.4	239464	2	AC103079	AC103079 Rattus no
18		176.8	16.4	267540	2	AC107153	AC107153 Rattus no
19		174.8	16.2	1238	9	BC025970	BC025970 Homo sapi
20		174.8	16.2	1330	6	AX714539	AX714539 Sequence
21		174.8	16.2	1330	9	AK056946	AK056946 Homo sapi
22		174.8	16.2	96875	9	AC008105	AC008105 Homo sapi
23		174.2	16.1	111803	2	AC136172	AC136172 Rattus no
24	c	173.2	16.0	2048	9	AK096785	AK096785 Homo sapi
25		172	15.9	223442	10	AL662804	AL662804 Mouse DNA
26		138	12.8	342	6	BD059439	BD059439 Secreted
27		110.8	10.3	225116	2	AC112538	AC112538 Rattus no
28		104.6	9.7	124088	10	AL929026	AL929026 Mouse DNA
29		104.4	9.7	240931	2	AC107097	AC107097 Rattus no
30	c	99.4	9.2	82400	10	AC090495	AC090495 Genomic s
31	c	96.4	8.9	171570	10	AL807807	AL807807 Mouse DNA
32	c	95.6	8.9	70748	2	AC137575	AC137575 Homo sapi
33		93.2	8.6	160539	2	AC109193	AC109193 Mus muscu
34	c	93	8.6	94058	2	AC141024	AC141024 Rattus no
35		91.6	8.5	76941	10	BX000698	BX000698 Mouse DNA
36		91.2	8.4	152435	10	AC121566	AC121566 Mus muscu
37	c	90.2	8.4	178757	10	AC122509	AC122509 Mus muscu
38		89.6	8.3	165592	10	AC122020	AC122020 Mus muscu
39	c	88.6	8.2	177100	2	AC112388	AC112388 Rattus no
40	c	88.6	8.2	246774	2	AC095078	AC095078 Rattus no
41	c	87.2	8.1	190944	2	AC113108	AC113108 Mus muscu
42		87.2	8.1	224198	2	AC106674	AC106674 Rattus no
43		87	8.1	205054	10	AC098712	AC098712 Mus muscu
44		86.6	8.0	201074	2	AC102097	AC102097 Mus muscu
45		86.2	8.0	178620	10	AC121865	AC121865 Mus muscu

ALIGNMENTS

RESULT 1
BD106410
LOCUS BD106410 2199 bp DNA linear PAT 18-SEP-2002
DEFINITION Secreted proteins and polynucleotides encoding them.
ACCESSION BD106410
VERSION BD106410.1 GI:23201228
KEYWORDS JP 2002503955-A/1.
SOURCE Chlamydia sp.
ORGANISM Chlamydia sp.
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE 1 (bases 1 to 2199)
AUTHORS Jacobs,K., Mccoy,J.M., Lavallie,E.R., Racie,L.A., Merberg,D., Treacy,M., Spaulding,V. and Agostino,M.J.
TITLE Secreted proteins and polynucleotides encoding them
JOURNAL Patent: JP 2002503955-A 1 05-FEB-2002;

```
            GENETICS INSTITUTE INC
COMMENT     PN   JP 2002503955-A/1
            PD   05-FEB-2002
            PF   20-MAR-1998 JP 1998545874
            PR   21-MAR-1997 US    08/822167,19-MAR-1998 US    09/044466 PI
            KENNETH JACOBS,JOHN M MCCOY,EDWARD R LAVALLIE,LISA A RACIE, PI
            DAVID MERBERG,
            PI   MAURICE TREACY,VIKKI SPAULDING,MICHAEL J AGOSTINO PC
            C12N15/12,C07K14/47,A61K38/17
            CC   Strandedness: Double;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers.
FEATURES            Location/Qualifiers
     source         1. .2199
                    /organism="Chlamydia sp."
                    /mol_type="genomic DNA"
                    /db_xref="taxon:35827"
BASE COUNT      552 a    511 c    674 g    462 t
ORIGIN

  Query Match           99.7%;  Score 1076.8;  DB 6;  Length 2199;
  Best Local Similarity  99.8%;  Pred. No. 4.6e-169;
  Matches 1078;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        707 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 766

Qy         61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        767 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 826

Qy        121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        827 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 886

Qy        181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        887 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 946

Qy        241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        947 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 1006

Qy        301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1007 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 1066

Qy        361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1067 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 1126

Qy        421 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAG 480
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       1127 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 1186

Qy        481 AAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAG 540
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       1187 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 1246

Qy        541 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1247 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 1306

Qy        601 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1307 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 1366

Qy        661 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1367 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 1426

Qy        721 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1427 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 1486

Qy        781 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 1546

Qy        841 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 1606

Qy        901 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 1666

Qy        961 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1726

Qy       1021 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1786


RESULT 2
AB021179
LOCUS       AB021179               3624 bp    mRNA    linear   PRI 05-DEC-2000
DEFINITION  Homo sapiens mRNA for HEXIM1 protein, complete cds.
ACCESSION   AB021179
VERSION     AB021179.1  GI:4062855
KEYWORDS    HEXIM1; HEXIM1 protein.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Kusuhara,M., Nagasaki,K., Kimura,K., Maass,N., Manabe,T.,
            Ishikawa,S., Aikawa,M., Miyazaki,K. and Yamaguchi,K.
  TITLE     Cloning of hexamethylene-bis-acetamide-inducible transcript,
            HEXIM1, in human vascular smooth muscle cells
  JOURNAL   Biomed. Res. 20, 273-279 (1999)
REFERENCE   2  (bases 1 to 3624)
  AUTHORS   Kusuhara,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-DEC-1998) Masatoshi Kusuhara, National Defense
            Medical College, The First Department of Internal Medicine; 3-2
            Namiki, Tokorozawa, Saitama 359-8513, Japan
            (E-mail:mkusu@me.ndmc.ac.jp, Tel:+81-42-995-1597,
            Fax:+81-42-996-5200)
FEATURES            Location/Qualifiers
     source         1. .3624
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /cell_type="coronary artery smooth muscle cell"
     gene           1. .3624
                    /gene="HEXIM1"
     CDS            690. .1769
                    /gene="HEXIM1"
                    /codon_start=1
                    /product="HEXIM1 protein"
                    /protein_id="BAA36166.1"
                    /db_xref="GI:4062856"
                    /translation="MAEPFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEED
                    SRWQSRAFPQLGGRPGPEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGD
                    FPPPAEVEPTPEAELLAQPCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKK
                    KRHWKPYYKLTWEEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEP
                    DLKTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHT
                    ESLQNMSKQELIKEYLELEKCLSRMEDENNRLRLESKRLGGDDARVRELELELDRLRA
                    ENLQLLTENELHRQQERAPLSKFGD"
BASE COUNT      978 a    775 c    964 g    907 t
ORIGIN
```

```
  Query Match           99.7%;  Score 1076.8;  DB 9;  Length 3624;
  Best Local Similarity  99.8%;  Pred. No. 4.3e-169;
  Matches 1078;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        690 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 749

Qy         61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        750 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 809

Qy        121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        810 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 869

Qy        181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        870 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 929

Qy        241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        930 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 989

Qy        301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        990 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 1049

Qy        361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1050 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 1109

Qy        421 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAG 480
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       1110 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 1169

Qy        481 AAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAG 540
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db       1170 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 1229

Qy        541 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1230 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 1289

Qy        601 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1290 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 1349

Qy        661 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1350 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 1409

Qy        721 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1410 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 1469

Qy        781 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1470 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 1529

Qy        841 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1530 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 1589

Qy        901 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1590 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 1649

Qy        961 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1650 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1709

Qy       1021 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1710 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1769
```

```
RESULT 3
AC142472/c
LOCUS       AC142472               35050 bp    DNA     linear   PRI 04-APR-2003
DEFINITION  Homo sapiens chromosome 17, clone XXfos-8119G3, complete sequence.
ACCESSION   AC142472
VERSION     AC142472.1  GI:29540490
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 35050)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone XXfos-8119G3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 35050)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-APR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L29477
                Center clone name: 8119_G_3
            --------------

            Only the last 35.1 kilobases of this clone are being submitted.
            The remainder overlaps accession number AC008105 [WICGR project
            L890].
FEATURES             Location/Qualifiers
     source          1. .35050
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="XXfos-8119G3"
                     /clone_lib="Human Female Fosmid Library"
     repeat_region   14. .320
                     /rpt_family="AluSp"
     repeat_region   1947. .1983
                     /rpt_family="AT_rich"
     repeat_region   3469. .3493
```

/rpt_family="(A)n"
repeat_region 3993. .4099
/rpt_family="C-rich"
repeat_region 5028. .5069
/rpt_family="G-rich"
repeat_region complement(5605. .5818)
/rpt_family="MIR"
repeat_region complement(6064. .6196)
/rpt_family="L2"
repeat_region complement(6713. .6862)
/rpt_family="MIR"
repeat_region 6863. .7166
/rpt_family="AluSc"
repeat_region complement(7167. .7225)
/rpt_family="MIR"
repeat_region 8048. .8104
/rpt_family="(CCCCG)n"
repeat_region 8113. .8149
/rpt_family="GC_rich"
repeat_region 9472. .9732
/rpt_family="L1MC4"
repeat_region 9733. .10036
/rpt_family="AluSg1"
repeat_region 10037. .10091
/rpt_family="L1MC4"
repeat_region complement(10169. .10477)
/rpt_family="AluSx"
repeat_region 10480. .10739
/rpt_family="L1MC4"
repeat_region 10750. .10906
/rpt_family="L1MC5"
repeat_region complement(10907. .11182)
/rpt_family="AluJo"
repeat_region 11183. .11552
/rpt_family="L1MC5"
repeat_region complement(11556. .11857)
/rpt_family="AluSx"
repeat_region complement(11913. .12044)
/rpt_family="FLAM_C"
repeat_region 12108. .12240
/rpt_family="FLAM_C"
repeat_region 12296. .12374
/rpt_family="AluS"
unsure 12353. .12373
/note="<30 qual SNGL region"
unsure 12374
/note="unresolved sequence region:
100 bp of sequence missing here"
repeat_region 12375. .12673
/rpt_family="AluSc"
repeat_region 13059. .13232
/rpt_family="MIR"
repeat_region complement(14041. .14336)
/rpt_family="AluSx"
repeat_region complement(14518. .14598)
/rpt_family="MIR"
repeat_region 17580. .17884
/rpt_family="AluSx"
repeat_region 17893. .18028
/rpt_family="L1MA3"
repeat_region complement(18032. .18350)
/rpt_family="AluY"
repeat_region 18360. .18383
/rpt_family="AT_rich"
repeat_region complement(18387. .18515)
/rpt_family="FLAM_C"
repeat_region 18533. .18707
/rpt_family="FRAM"
repeat_region 18708. .18877
/rpt_family="L1MA4"
repeat_region 19015. .19324
/rpt_family="AluYa5"
repeat_region 20221. .20288
/rpt_family="(CCCCG)n"
repeat_region 20403. .20429
/rpt_family="GC_rich"
repeat_region complement(21264. .21365)
/rpt_family="MIR"
repeat_region complement(21662. .21734)
/rpt_family="L2"
repeat_region complement(22401. .22550)
/rpt_family="L2"
repeat_region complement(23139. .23338)
/rpt_family="MIR"
repeat_region complement(23346. .23673)
/rpt_family="AluJo"
repeat_region complement(23704. .23903)
/rpt_family="MIR"
repeat_region complement(23996. .24125)
/rpt_family="MER53"
repeat_region 24057. .24138
/rpt_family="MER53"
repeat_region 25014. .25095
/rpt_family="(CA)n"
repeat_region 25274. .25344
/rpt_family="MIR3"
repeat_region complement(25687. .25742)
/rpt_family="L2"
repeat_region complement(25847. .25917)
/rpt_family="L1MB8"
repeat_region 25918. .26185
/rpt_family="AluJo"
repeat_region complement(26186. .26209)
/rpt_family="L1MB8"
repeat_region 26243. .26497
/rpt_family="AluYc"
repeat_region 26500. .26799
/rpt_family="AluY"
repeat_region complement(26800. .27209)
/rpt_family="L1MB4"
repeat_region complement(27220. .27528)
/rpt_family="AluSx"
repeat_region complement(27533. .27592)
/rpt_family="AluSq/x"
repeat_region complement(27597. .27800)
/rpt_family="AluSx"
repeat_region complement(27805. .27916)
/rpt_family="AluSq/x"
repeat_region complement(27917. .27963)
/rpt_family="L1MB4"
repeat_region 29143. .29261
/rpt_family="MER5B"
repeat_region complement(29547. .29628)
/rpt_family="L2"
repeat_region complement(30172. .30304)
/rpt_family="MIR"
repeat_region 30327. .30450
/rpt_family="MIR"
repeat_region complement(30485. .30611)
/rpt_family="AluSg/x"
repeat_region 30675. .30825
/rpt_family="L3b"
repeat_region complement(30871. .31136)

Query Match 99.7%; Score 1076.8; DB 9; Length 35050;
Best Local Similarity 99.8%; Pred. No. 3.1e-169;
Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3397 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 3338

Qy 61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3337 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 3278

```
Qy        121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3277 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 3218

Qy        181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3217 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 3158

Qy        241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3157 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 3098

Qy        301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3097 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 3038

Qy        361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3037 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 2978

Qy        421 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAG 480
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       2977 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 2918

Qy        481 AAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAG 540
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       2917 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 2858

Qy        541 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2857 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 2798

Qy        601 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2797 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 2738

Qy        661 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2737 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 2678

Qy        721 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2677 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 2618

Qy        781 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2617 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 2558

Qy        841 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2557 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 2498

Qy        901 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2497 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 2438

Qy        961 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2437 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 2378

Qy       1021 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2377 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 2318
```

RESULT 4
AC138150
LOCUS AC138150 38849 bp DNA linear PRI 14-FEB-2003
DEFINITION Homo sapiens chromosome 17, clone RP13-890H12, complete sequence.
ACCESSION AC138150
VERSION AC138150.4 GI:28376773
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 38849)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone RP13-890H12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 38849)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (18-DEC-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 38849)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (31-JAN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 4 (bases 1 to 38849)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,

```
          Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL Submitted (14-FEB-2003) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Feb 14, 2003 this sequence version replaced gi:28173166.
          All repeats were identified using RepeatMasker:
          Smit, A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          -------------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
          -------------- Project Information
              Center project name: L28731
              Center clone name: 890_H_12
          --------------.
FEATURES             Location/Qualifiers
     source          1. .38849
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP13-890H12"
                     /clone_lib="RPCI-13 Human Female BAC"
     repeat_region   84. .385
                     /rpt_family="AluSx"
     repeat_region   complement(389. .758)
                     /rpt_family="L1MC5"
     repeat_region   759. .1034
                     /rpt_family="AluJo"
     repeat_region   complement(1035. .1191)
                     /rpt_family="L1MC5"
     repeat_region   complement(1202. .1461)
                     /rpt_family="L1MC4"
     repeat_region   1464. .1772
                     /rpt_family="AluSx"
     repeat_region   complement(1850. .1904)
                     /rpt_family="L1MC4"
     repeat_region   complement(1905. .2208)
                     /rpt_family="AluSg1"
     repeat_region   complement(2209. .2469)
                     /rpt_family="L1MC4"
     repeat_region   3792. .3828
                     /rpt_family="GC_rich"
     repeat_region   3837. .3893
                     /rpt_family="(CGGGG)n"
     repeat_region   4716. .4774
                     /rpt_family="MIR"
     repeat_region   complement(4775. .5078)
                     /rpt_family="AluSc"
     repeat_region   5079. .5204
                     /rpt_family="MIR"
     repeat_region   5745. .5877
                     /rpt_family="L2"
     repeat_region   6123. .6336
                     /rpt_family="MIR3"
     repeat_region   6872. .6913
                     /rpt_family="C-rich"
     repeat_region   7842. .7948
                     /rpt_family="G-rich"
     repeat_region   8448. .8472
                     /rpt_family="(T)n"
     repeat_region   9962. .9998
                     /rpt_family="AT_rich"
     repeat_region   complement(11625. .11932)
                     /rpt_family="AluSp"
     repeat_region   12103. .12389
                     /rpt_family="AluSp"
     repeat_region   complement(12567. .12862)
                     /rpt_family="AluSx"
     repeat_region   complement(12863. .13169)
                     /rpt_family="AluJb"
     repeat_region   complement(13209. .13491)
                     /rpt_family="AluY"
     repeat_region   13566. .13589
                     /rpt_family="AT_rich"
     repeat_region   13595. .13903
                     /rpt_family="AluSp"
     repeat_region   13904. .14038
                     /rpt_family="AluSx"
     repeat_region   14039. .14352
                     /rpt_family="AluY"
     repeat_region   14353. .14522
                     /rpt_family="AluSx"
     repeat_region   14558. .14685
                     /rpt_family="FLAM_C"
     repeat_region   complement(14704. .14854)
                     /rpt_family="FRAM"
     repeat_region   15205. .15307
                     /rpt_family="AluJb"
     repeat_region   15359. .15633
                     /rpt_family="AluSx"
     repeat_region   16074. .16377
                     /rpt_family="AluSc"
     repeat_region   16380. .16558
                     /rpt_family="AluSg/x"
     repeat_region   16743. .16877
                     /rpt_family="AluSx"
     repeat_region   16878. .17181
                     /rpt_family="AluSp"
     repeat_region   17182. .17348
                     /rpt_family="AluSx"
     repeat_region   complement(18100. .18399)
                     /rpt_family="AluSc"
     repeat_region   complement(18948. .19238)
                     /rpt_family="AluSx"
     repeat_region   22400. .22707
                     /rpt_family="AluSx"
     repeat_region   complement(22708. .22845)
                     /rpt_family="MIR"
     repeat_region   22960. .23228
                     /rpt_family="AluSx"
     repeat_region   23231. .23524
                     /rpt_family="AluSq"
     repeat_region   23532. .23838
                     /rpt_family="AluSx"
     repeat_region   23843. .24006
                     /rpt_family="AluSg/x"
     repeat_region   25108. .25151
```

```
  Query Match             99.7%;  Score 1076.8;  DB 9;  Length 38849;
  Best Local Similarity   99.8%;  Pred. No. 3.1e-169;
  Matches 1078;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8544 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 8603

Qy         61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8604 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 8663

Qy        121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8664 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 8723

Qy        181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8724 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 8783

Qy        241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8784 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 8843
```

```
Qy        301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8844 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 8903

Qy        361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8904 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 8963

Qy        421 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAG 480
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       8964 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 9023

Qy        481 AAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAG 540
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       9024 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 9083

Qy        541 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9084 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 9143

Qy        601 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9144 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 9203

Qy        661 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9204 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 9263

Qy        721 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9264 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 9323

Qy        781 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9324 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 9383

Qy        841 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9384 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 9443

Qy        901 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9444 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 9503

Qy        961 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9504 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 9563

Qy       1021 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9564 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 9623
```

```
RESULT 5
AC087298/c
LOCUS       AC087298              100836 bp    DNA     linear   HTG 20-APR-2002
DEFINITION  Homo sapiens chromosome 17 clone RP11-403G3 map 17, WORKING DRAFT
            SEQUENCE, 15 ordered pieces.
ACCESSION   AC087298
VERSION     AC087298.8  GI:20219131
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 100836)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-403G3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 100836)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-DEC-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 100836)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 20, 2002 this sequence version replaced gi:20177719.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12020
                Center clone name: 403_G_3
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 96277 bases at least Q40
                Consensus quality: 98051 bases at least Q30
                Consensus quality: 98838 bases at least Q20
                Insert size: 140000; agarose-fp
                Insert size: 99436; sum-of-contigs
                Quality coverage: 19.1 in Q20 bases; agarose-fp
                Quality coverage: 26.9 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. Gaps between the contigs
```

```
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *       1     8098: contig of 8098 bp in length
          *    8099     8198: gap of 100 bp
          *    8199     8271: contig of 73 bp in length
          *    8272     8371: gap of 100 bp
          *    8372     9497: contig of 1126 bp in length
          *    9498     9597: gap of 100 bp
          *    9598    10863: contig of 1266 bp in length
          *   10864    10963: gap of 100 bp
          *   10964    12232: contig of 1269 bp in length
          *   12233    12332: gap of 100 bp
          *   12333    13844: contig of 1512 bp in length
          *   13845    13944: gap of 100 bp
          *   13945    15215: contig of 1271 bp in length
          *   15216    15315: gap of 100 bp
          *   15316    18234: contig of 2919 bp in length
          *   18235    18334: gap of 100 bp
          *   18335    21787: contig of 3453 bp in length
          *   21788    21887: gap of 100 bp
          *   21888    25485: contig of 3598 bp in length
          *   25486    25585: gap of 100 bp
          *   25586    40531: contig of 14946 bp in length
          *   40532    40631: gap of 100 bp
          *   40632    52119: contig of 11488 bp in length
          *   52120    52219: gap of 100 bp
          *   52220    64941: contig of 12722 bp in length
          *   64942    65041: gap of 100 bp
          *   65042    79587: contig of 14546 bp in length
          *   79588    79687: gap of 100 bp
          *   79688   100836: contig of 21149 bp in length.
FEATURES             Location/Qualifiers
     source          1. .100836
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-403G3"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .8098
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    8199. .8271
                     /note="assembly_fragment"
     misc_feature    8372. .9497
                     /note="assembly_fragment"
     misc_feature    9598. .10863
                     /note="assembly_fragment"
     misc_feature    10964. .12232
                     /note="assembly_fragment"
     misc_feature    12333. .13844
                     /note="assembly_fragment"
     misc_feature    13945. .15215
                     /note="assembly_fragment"
     misc_feature    15316. .18234
                     /note="assembly_fragment"
     misc_feature    18335. .21787
                     /note="assembly_fragment"
     misc_feature    21888. .25485
                     /note="assembly_fragment"
     misc_feature    25586. .40531
                     /note="assembly_fragment"
     misc_feature    40632. .52119
                     /note="assembly_fragment"
     misc_feature    52220. .64941
                     /note="assembly_fragment"
     misc_feature    65042. .79587
                     /note="assembly_fragment"
     misc_feature    79688. .100836
                     /note="assembly_fragment"
BASE COUNT    23052 a  27082 c  27053 g  22249 t   1400 others
ORIGIN

  Query Match             99.7%;  Score 1076.8;  DB 2;  Length 100836;
  Best Local Similarity   99.8%;  Pred. No. 2.7e-169;
  Matches 1078;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70785 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 70726

Qy         61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70725 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 70666

Qy        121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70665 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 70606

Qy        181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70605 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 70546

Qy        241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70545 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 70486

Qy        301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70485 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 70426

Qy        361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70425 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 70366

Qy        421 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAG 480
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      70365 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 70306

Qy        481 AAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAG 540
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      70305 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 70246

Qy        541 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70245 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 70186

Qy        601 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70185 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 70126

Qy        661 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70125 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 70066

Qy        721 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70065 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 70006

Qy        781 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70005 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 69946

Qy        841 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69945 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 69886

Qy        901 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      69885 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 69826

Qy        961 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69825 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 69766

Qy       1021 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69765 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 69706


RESULT 6
BD160090
LOCUS       BD160090               2178 bp    DNA     linear   PAT 17-JAN-2003
DEFINITION  Primer for synthesizing full-length cDNA and use thereof.
ACCESSION   BD160090
VERSION     BD160090.1  GI:27865848
KEYWORDS    JP 2002191363-A/14933.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2178)
  AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
            Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
  TITLE     Primer for synthesizing full-length cDNA and use thereof
  JOURNAL   Patent: JP 2002191363-A 14933 09-JUL-2002;
            HELIX RESEARCH INSTITUTE
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002191363-A/14933
            PD   09-JUL-2002
            PF   28-JUL-2000 JP 2000280990
            PI   TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
            PI   SAITO,
            PI   JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
            PI   KEIICHI NAGAI,TETSUJI OTSUKI
            PC
            C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
             10,
            PC   C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
             Primer for synthesizing full-length cDNA and use thereof FH   Key
                      Location/Qualifiers
            FT   CDS             (716). .(1792).
FEATURES             Location/Qualifiers
     source          1. .2178
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      525 a    513 c    678 g    462 t
ORIGIN

  Query Match           99.6%;  Score 1075.2;  DB 6;  Length 2178;
  Best Local Similarity   99.7%;  Pred. No. 8.4e-169;
  Matches 1077;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        716 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 775

Qy         61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        776 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 835

Qy        121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        836 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 895

Qy        181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        896 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 955

Qy        241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        956 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 1015

Qy        301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1016 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 1075

Qy        361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1076 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 1135

Qy        421 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAG 480
              |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1136 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 1195

Qy        481 AAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAG 540
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       1196 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 1255

Qy        541 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1256 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 1315

Qy        601 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1316 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 1375

Qy        661 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1376 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 1435

Qy        721 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1436 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 1495

Qy        781 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1496 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 1555

Qy        841 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1556 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 1615

Qy        901 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1616 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 1675

Qy        961 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1676 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1735

Qy       1021 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080
              |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1736 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCCTTCCAAGTTTGGAGACTAG 1795


RESULT 7
AK023624
LOCUS       AK023624               2178 bp    mRNA    linear   PRI 01-AUG-2002
DEFINITION  Homo sapiens cDNA FLJ13562 fis, clone PLACE1008080, highly similar
            to Homo sapiens mRNA for HEXIM1 protein.
ACCESSION   AK023624
VERSION     AK023624.1  GI:10435606
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
            Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
            Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
            Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
```

```
            Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
            Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2178)
  AUTHORS   Isogai,T. and Otsuki,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
            Helix Research Institute (supported by Japan Key Technology Center
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES             Location/Qualifiers
     source          1. .2178
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="PLACE1008080"
                     /tissue_type="placenta"
                     /clone_lib="PLACE1"
                     /note="cloning vector: pME18SFL3"
BASE COUNT      525 a    513 c    678 g    462 t
ORIGIN
```

Query Match 99.6%; Score 1075.2; DB 9; Length 2178;
Best Local Similarity 99.7%; Pred. No. 8.4e-169;
Matches 1077; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy          1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        716 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 775

Qy         61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        776 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 835

Qy        121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        836 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 895

Qy        181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        896 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 955

Qy        241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        956 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 1015

Qy        301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1016 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 1075

Qy        361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1076 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 1135

Qy        421 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAG 480
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       1136 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 1195

Qy        481 AAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAG 540
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1196 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 1255

Qy        541 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1256 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 1315

Qy        601 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1316 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 1375

Qy        661 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1376 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 1435

Qy        721 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1436 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 1495

Qy        781 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1496 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 1555

Qy        841 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1556 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 1615

Qy        901 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1616 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 1675

Qy        961 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1676 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1735

Qy       1021 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080
              ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1736 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCCTTCCAAGTTTGGAGACTAG 1795
```

```
RESULT 8
BC006460
LOCUS       BC006460                2086 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, HMBA-inducible, clone MGC:1880 IMAGE:3535529, mRNA,
            complete cds.
ACCESSION   BC006460
VERSION     BC006460.1  GI:13623668
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2086)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-APR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 8 Row: d Column: 1
```

```
          This clone was selected for full length sequencing because it
          passed the following selection criteria: matched mRNA gi: 4062855.
FEATURES             Location/Qualifiers
     source          1. .2086
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="LocusID:10614"
                     /db_xref="taxon:9606"
                     /clone="MGC:1880 IMAGE:3535529"
                     /tissue_type="Lung, small cell carcinoma"
                     /clone_lib="NIH_MGC_7"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             606. .1685
                     /codon_start=1
                     /product="HMBA-inducible"
                     /protein_id="AAH06460.1"
                     /db_xref="GI:13623669"
                     /translation="MAEPFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEED
                     SRWQSRAFPQLGGRPGPEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGD
                     FPPPAEVEPTPEAELLAQPCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKK
                     KRHWKPYYKLTWEEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEP
                     DLKTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHT
                     ESLQNMSKQELIKEYLELEKCLSRMEDENNRLRLESKRLGGDDARVRELELELDRLRA
                     ENLQLLTENELHRQQERAPLSKFGD"
BASE COUNT      510 a    505 c    618 g    453 t
ORIGIN

  Query Match           99.4%;  Score 1073.6;  DB 9;  Length 2086;
  Best Local Similarity   99.6%;  Pred. No. 1.6e-168;
  Matches 1076;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        606 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 665

Qy         61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 725

Qy        121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        726 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 785

Qy        181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        786 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 845

Qy        241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 905

Qy        301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        906 GGCGACTTCCCGCCGCCGGCAGAAGTGGAGCCGACGCCCGAGGCCGAGCTGCTCGCCCAG 965

Qy        361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 1025

Qy        421 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAG 480
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1026 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 1085

Qy        481 AAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAG 540
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1086 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 1145

Qy        541 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 1205

Qy        601 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       1206 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGACCTC 1265

Qy        661 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1266 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 1325

Qy        721 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1326 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 1385

Qy        781 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1386 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 1445

Qy        841 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1446 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 1505

Qy        901 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1506 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 1565

Qy        961 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1566 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1625

Qy       1021 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1626 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1685


RESULT 9
AC024047/c
LOCUS       AC024047               142326 bp    DNA     linear   HTG 07-JUL-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-403G3, WORKING DRAFT
            SEQUENCE, 28 unordered pieces.
ACCESSION   AC024047
VERSION     AC024047.3  GI:8570008
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 142326)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 142326)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 16, 2000 this sequence version replaced gi:7109599.

            ---------------- Genome Center ----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0403G03
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 127942 bases at least Q40
            Consensus quality: 133308 bases at least Q30
            Consensus quality: 135734 bases at least Q20
            Insert size: 147000; agarose-fp
            Insert size: 139626; sum-of-contigs
```

```
Quality coverage: 4.06 in Q20 bases; agarose-fp
Quality coverage: 4.33 in Q20 bases; sum-of-contigs
-------------------------------------------------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1575: contig of 1575 bp in length
*     1576     1675: gap of unknown length
*     1676     3114: contig of 1439 bp in length
*     3115     3214: gap of unknown length
*     3215     5389: contig of 2175 bp in length
*     5390     5489: gap of unknown length
*     5490     7129: contig of 1640 bp in length
*     7130     7229: gap of unknown length
*     7230     9240: contig of 2011 bp in length
*     9241     9340: gap of unknown length
*     9341    11177: contig of 1837 bp in length
*    11178    11277: gap of unknown length
*    11278    13113: contig of 1836 bp in length
*    13114    13213: gap of unknown length
*    13214    16526: contig of 3313 bp in length
*    16527    16626: gap of unknown length
*    16627    19864: contig of 3238 bp in length
*    19865    19964: gap of unknown length
*    19965    22667: contig of 2703 bp in length
*    22668    22767: gap of unknown length
*    22768    26650: contig of 3883 bp in length
*    26651    26750: gap of unknown length
*    26751    29303: contig of 2553 bp in length
*    29304    29403: gap of unknown length
*    29404    32073: contig of 2670 bp in length
*    32074    32173: gap of unknown length
*    32174    37141: contig of 4968 bp in length
*    37142    37241: gap of unknown length
*    37242    40738: contig of 3497 bp in length
*    40739    40838: gap of unknown length
*    40839    45166: contig of 4328 bp in length
*    45167    45266: gap of unknown length
*    45267    49501: contig of 4235 bp in length
*    49502    49601: gap of unknown length
*    49602    55647: contig of 6046 bp in length
*    55648    55747: gap of unknown length
*    55748    61532: contig of 5785 bp in length
*    61533    61632: gap of unknown length
*    61633    68005: contig of 6373 bp in length
*    68006    68105: gap of unknown length
*    68106    75211: contig of 7106 bp in length
*    75212    75311: gap of unknown length
*    75312    81560: contig of 6249 bp in length
*    81561    81660: gap of unknown length
*    81661    88524: contig of 6864 bp in length
*    88525    88624: gap of unknown length
*    88625    95519: contig of 6895 bp in length
*    95520    95619: gap of unknown length
*    95620   105148: contig of 9529 bp in length
*   105149   105248: gap of unknown length
*   105249   115269: contig of 10021 bp in length
*   115270   115369: gap of unknown length
*   115370   128646: contig of 13277 bp in length
*   128647   128746: gap of unknown length
*   128747   142326: contig of 13580 bp in length.
FEATURES             Location/Qualifiers
     source          1. .142326
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /clone="RP11-403G3"
```

```
BASE COUNT    32503 a  37098 c  37256 g  32736 t   2733 others
ORIGIN

  Query Match            97.5%;  Score 1052.8;  DB 2;  Length 142326;
  Best Local Similarity  99.6%;  Pred. No. 2.5e-165;
  Matches 1076;  Conservative    0;  Mismatches    2;  Indels    2;  Gaps    2;

Qy          1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     108505 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 108446

Qy         61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     108445 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 108386

Qy        121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     108385 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 108326

Qy        181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     108325 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 108266

Qy        241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     108265 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 108206

Qy        301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db     108205 -GCGACTTCCCG-CGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 108148

Qy        361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     108147 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 108088

Qy        421 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAG 480
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     108087 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 108028

Qy        481 AAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAG 540
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db     108027 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 107968

Qy        541 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     107967 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 107908

Qy        601 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     107907 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 107848

Qy        661 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     107847 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 107788

Qy        721 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     107787 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 107728

Qy        781 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     107727 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 107668

Qy        841 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     107667 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 107608

Qy        901 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     107607 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 107548

Qy        961 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     107547 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 107488

Qy       1021 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     107487 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 107428


RESULT 10
AY090614
LOCUS       AY090614               3402 bp    mRNA    linear   ROD 18-JUL-2002
DEFINITION  Mus musculus cardiac lineage protein 1 (Clp1) mRNA, complete cds.
ACCESSION   AY090614
VERSION     AY090614.1  GI:20135640
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3402)
  AUTHORS   Huang,F., Wagner,M. and Siddiqui,M.A.Q.
  TITLE     Structure, expression, and functional characterization of the mouse
            CLP-1 gene
  JOURNAL   Gene 292 (1-2), 245-259 (2002)
  MEDLINE   22114986
   PUBMED   12119119
REFERENCE   2  (bases 1 to 3402)
  AUTHORS   Huang,F., Wagner,M. and Siddiqui,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2002) Anatomy and Cell Biology, State University
            of New York Health Science Center, 450 Clarkson Avenue, Brooklyn,
            NY 11203, USA
FEATURES             Location/Qualifiers
     source          1. .3402
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="129/SvJ"
                     /db_xref="taxon:10090"
     gene            1. .3402
                     /gene="Clp1"
     CDS             596. .1666
                     /gene="Clp1"
                     /note="nuclear protein; CLP-1"
                     /codon_start=1
                     /product="cardiac lineage protein 1"
                     /protein_id="AAM09026.1"
                     /db_xref="GI:20135641"
                     /translation="MAEPLLTEHQHQPQTSNCTGAAVVHEEHTSERPPSAEERVPKED
                     SRWQSRASLQSGSRPGQEGEGGLKHQLPPLQTNACPELSSLEKGEKGQNGEDLSTGGA
                     SPSAEGEPMSESLVQPGHDSEATKQEAPAAGGEEPWGQQQRQLGKKKHRRRPSKKKRH
                     WKPYYKLTWEEKKKFDEKQSLRASRVRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLK
                     TGLYPKRAAAKSDDTSDEDFVEEAGEEDGGSDGMGGDGSEFLQRDFSETYERYHAESL
                     QNMSKQELIKEYLELEKCLSRKEDENNRLRLESKRLGGVDARVRELELELDRLRAENL
                     QLLTENELHRQQERAPLSKFGD"
BASE COUNT      920 a    771 c    890 g    821 t
ORIGIN

  Query Match           77.9%;  Score 840.8;  DB 10;  Length 3402;
  Best Local Similarity  86.9%;  Pred. No. 6.1e-130;
  Matches  939;  Conservative    0;  Mismatches  132;  Indels    9;  Gaps    1;

Qy          1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60
              ||||||||||||| ||||| ||||| ||||||||||||||||||||||||||||||||||||
Db        596 ATGGCCGAGCCACTCTTGACAGAACATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 655

Qy         61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120
              ||||||| |||||| |||||||  | | ||||||||||||||| |||||||||| ||||||
Db        656 GCTGCTGTTGTCCATGAAGAGCATACCTCTGAGCGCCCCCCAAGCGCGGAGGAACGGGTG 715

Qy        121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180
              ||| ||||||||||||||||||||||||||||| |   ||||  ||| ||||||| |||
Db        716 CCCAAGGAGGACAGTAGGTGGCAATCGAGAGCGTCCTTGCAGTCCGGTAGCCGTCCAGGG 775

Qy        181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240
              | |||||| |||||| ||||| |   ||| |  || || ||||||||  | |||||||||
Db        776 CAGGAGGGAGAAGGGGGCCTGAAGCACCAGCTGCCCCCATTGCAGACGAATGCCTGTCCA 835

Qy        241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300
              ||||  |||| ||||    |||||||||||||||||||||||||| ||||  ||| |||||
Db        836 GAATTGAGCTCCCTGGAGAAGGGCGAGAAGGGCCAGAATGGGGAGGACTTATCCACTGGC 895

Qy        301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360
              || | || |||| || |||  | ||           |||   | | |||   |||| |||
Db        896 GGTGCCTCCCCGTCGGCGGAGGGAGA---------GCCGATGTCAGAGTCCCTCGTGCAG 946

Qy        361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 420
              ||  |||||||||| |||||||  |||  || ||||||||||| || ||||| |||
Db        947 CCAGGTCATGACTCGGAGGCCACCAAGCAGGAGGCTCCTGCCGCCGGAGGCGAGGAGCCA 1006

Qy        421 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAG 480
              |||||||||| |||||||||||||| |||||||||||  ||||||||| || ||||||
Db       1007 TGGGGACAGCAACAGAGACAGCTGGGCAAGAAAAAACATCGGAGACGCCCCTCAAAGAAG 1066

Qy        481 AAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAG 540
              |||||||||||||| || |||||||||||  ||||| ||||||||||||||||||||||
Db       1067 AAGCGGCATTGGAAGCCCTACTACAAGCTGACTTGGGAGGAGAAGAAAAAGTTCGACGAG 1126

Qy        541 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 600
              || |||||||| ||||||||  || | ||||||||||||||||||||||||||||||||||
Db       1127 AAGCAGAGCCTGCGAGCTTCGCGGGTTCGAGCCGAGATGTTCGCCAAGGGCCAGCCAGTT 1186

Qy        601 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 660
              |||||||||||||||||||||||||||||||||||||||||| ||||| ||||||||||
Db       1187 GCGCCCTATAACACCACGCAGTTCCTCATGGATGACCACGATCAGGAGGAGCCTGATCTC 1246

Qy        661 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 720
              |||||||||| ||| |||||||||| ||||||||||||||||||||||||||||||||||
Db       1247 AAAACCGGCCTTTACCCCAAGCGGGCAGCCGCCAAATCCGACGACACCAGCGATGAGGAT 1306

Qy        721 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 780
              ||  |||||||||  |||||||| || || |||||||||| ||||||||||||||||||
Db       1307 TTTGTGGAAGAAGCTGGTGAGGAAGACGGAGGCAGCGATGGCATGGGAGGGGACGGCAGC 1366

Qy        781 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 840
              |||||||||||||||||||||||||||||||||||||||||||| | ||||||||||||
Db       1367 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACGCCGAGAGCCTGCAG 1426

Qy        841 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 900
              |||||||||||||||||||||||| |||||||||| |||||||||||||||| |||
Db       1427 AACATGAGCAAGCAGGAGCTCATCAAAGAGTACCTGGAGCTGGAGAAGTGCCTCTCCCGC 1486

Qy        901 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 960
              | ||||||||| || |||||||||||||||| |||||||||||| |||| |||||||
Db       1487 AAGGAGGACGAAAATAACCGGCTGCGGCTGGAAAGCAAGCGGCTGGGAGGCGTCGACGCG 1546

Qy        961 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1020
              || |||||||||| |||| |||||||||||||||||||||||||||| |||||||||||||
Db       1547 CGAGTGCGGGAGCTCGAGCTAGAGCTGGACCGGCTGCGCGCTGAGAACCTCCAGCTGCTG 1606

Qy       1021 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080
              |||||||||||||||||||||||||||||||||||||||||| |||||||||| ||||||
Db       1607 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCTCTTTCCAAGTTCGGCGACTAG 1666


RESULT 11
AL731805
LOCUS       AL731805               111002 bp    DNA     linear   ROD 21-JUN-2002
DEFINITION  Mouse DNA sequence from clone RP23-358E19 on chromosome 11,
            complete sequence.
ACCESSION   AL731805
VERSION     AL731805.8  GI:21540125
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 111002)
```

```
  AUTHORS   Dunn,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 23, 2002 this sequence version replaced gi:21531483.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-358E19 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES              Location/Qualifiers
     source           1. .111002
                      /organism="Mus musculus"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:10090"
                      /chromosome="11"
                      /clone="RP23-358E19"
                      /clone_lib="RPCI-23"
BASE COUNT    28051 a  27579 c  27135 g  28237 t
ORIGIN

  Query Match           77.9%;  Score 840.8;  DB 10;  Length 111002;
  Best Local Similarity  86.9%;  Pred. No. 3.8e-130;
  Matches  939;  Conservative    0;  Mismatches  132;  Indels    9;  Gaps    1;

Qy          1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60
              |||||||||||| |||| |||| ||||||||||||||||||||||||||||||||||||
Db      91440 ATGGCCGAGCCACTCTTGACAGAACATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 91499

Qy         61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120
              ||||||| |||||| |||||||  | | |||||||||||||| |||||||||| ||||||
Db      91500 GCTGCTGTTGTCCATGAAGAGCATACCTCTGAGCGCCCCCCCAAGCGCGGAGGAACGGGTG 91559

Qy        121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180
              ||| |||||||||||||||||||||||||||||| |   |||| ||| ||||||| |||
Db      91560 CCCAAGGAGGACAGTAGGTGGCAATCGAGAGCGTCCTTGCAGTCCGGTAGCCGTCCAGGG 91619

Qy        181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240
              | |||||| |||||| ||||| |   ||| |  || || ||||||||  | |||||||||
Db      91620 CAGGAGGGAGAAGGGGGCCTGAAGCACCAGCTGCCCCCATTGCAGACGAATGCCTGTCCA 91679

Qy        241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300
              ||||  |||| ||||    |||||||||||||||||||||||| |||| ||| |||||
Db      91680 GAATTGAGCTCCCTGGAGAAGGGCGAGAAGGGCCAGAATGGGGAGGACTTATCCACTGGC 91739

Qy        301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360
              || | || |||| || ||| | ||          |||  | | |||  |||| |||
Db.     91740 GGTGCCTCCCCGTCGGCGGAGGGAGA---------GCCGATGTCAGAGTCCCTCGTGCAG 91790

Qy        361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 420
              ||  ||||||||| |||||||  |||  || |||||||||||| || ||||| |||
Db      91791 CCAGGTCATGACTCGGAGGCCACCAAGCAGGAGGCTCCTGCCGCCGGAGGCGAGGAGCCA 91850

Qy        421 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAG 480
              |||||||||| |||||||||||||| |||||||||||  ||||||||| || ||||||
Db      91851 TGGGGACAGCAACAGAGACAGCTGGGCAAGAAAAAACATCGGAGACGCCCCTCAAAGAAG 91910

Qy        481 AAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAG 540
              |||||||||||||| || ||||||||||||  ||||| ||||||||||||||||||||||
Db      91911 AAGCGGCATTGGAAGCCCTACTACAAGCTGACTTGGGAGGAGAAGAAAAAGTTCGACGAG 91970

Qy        541 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 600
              || |||||||| ||||||||  || | ||||||||||||||||||||||||||||||||
Db      91971 AAGCAGAGCCTGCGAGCTTCGCGGGTTCGAGCCGAGATGTTCGCCAAGGGCCAGCCAGTT 92030

Qy        601 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 660
              |||||||||||||||||||||||||||||||||||||| ||||| |||||||||| ||||||
Db      92031 GCGCCCTATAACACCACGCAGTTCCTCATGGATGACCACGATCAGGAGGAGCCTGATCTC 92090

Qy        661 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 720
              |||||||||| ||| |||||||||| ||||||||||||||||||||||||||||||||||
Db      92091 AAAACCGGCCTTTACCCCAAGCGGGCAGCCGCCAAATCCGACGACACCAGCGATGAGGAT 92150

Qy        721 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 780
              ||  |||||||||  |||||||| || || |||||||||| ||||||||||||||||||
Db      92151 TTTGTGGAAGAAGCTGGTGAGGAAGACGGAGGCAGCGATGGCATGGGAGGGGACGGCAGC 92210

Qy        781 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 840
              |||||||||||||||||||||||||||||||||||||||||||||||| | ||||||||||||
Db      92211 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACGCCGAGAGCCTGCAG 92270

Qy        841 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 900
              |||||||||||||||||||||||||||||||| |||||||||| ||||||||||||||||
Db      92271 AACATGAGCAAGCAGGAGCTCATCAAAGAGTACCTGGAGCTGGAGAAGTGCCTCTCCCGC 92330

Qy        901 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 960
              | ||||||||| || |||||||||||||||| |||||||||||||| |||| |||||||
Db      92331 AAGGAGGACGAAAATAACCGGCTGCGGCTGGAAAGCAAGCGGCTGGGAGGCGTCGACGCG 92390

Qy        961 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1020
              || |||||||||| ||||| |||||||||||||||||||| ||||||||||||||||||
Db      92391 CGAGTGCGGGAGCTCGAGCTAGAGCTGGACCGGCTGCGCGCTGAGAACCTCCAGCTGCTG 92450

Qy       1021 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080
              |||||||||||||||||||||||||||||||||||||||||| |||||||||| ||||||
Db      92451 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCTCTTTCCAAGTTCGGCGACTAG 92510

RESULT 12
BC022111
LOCUS       BC022111                1488 bp    mRNA    linear   ROD 16-APR-2003
DEFINITION  Mus musculus cardiac lineage protein 1, mRNA (cDNA clone MGC:36249
            IMAGE:5067620), complete cds.
ACCESSION   BC022111
VERSION     BC022111.1  GI:18381039
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1488)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
```

Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

MEDLINE 22388257

PUBMED 12477932

REFERENCE 2 (bases 1 to 1488)

AUTHORS Strausberg,R.

TITLE Direct Submission

JOURNAL Submitted (24-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov

COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 57 Row: a Column: 15
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.

FEATURES Location/Qualifiers

source 1. .1488
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="MGC:36249 IMAGE:5067620"
/tissue_type="Mammary tumor metastatized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."
/clone_lib="NCI_CGAP_Lu29"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

gene 1. .1488
/gene="Clp1-pending"
/note="synonyms: HIS1, CLP-1"
/db_xref="LocusID:192231"
/db_xref="MGI:2385923"

CDS 448. .1188
/codon_start=1
/product="Clp1-pending protein"
/protein_id="AAH22111.1"
/db_xref="GI:18381040"
/db_xref="LocusID:192231"
/translation="MSESLVQPGHDSEATKQEAPAAGGEEPWGQQQRQLGKKKHRRRP SKKKRHWKPYYKLTWEEKKKFDEKQSLRASRVRAEMFAKGQPVAPYNTTQFLMDDHDQ EEPDLKTGLYPKRAAAKSDDTSDEDFVEEAGEEDGGSDGMGGDGSEFLQRDFSETYER YHAESLQNMSKQELIKEYLELEKCLSRKEDENNRLRLESKRLGGVDARVRELELELDR LRAENLQLLTENELHRQQERAPLSKFGD"

BASE COUNT 385 a 371 c 449 g 283 t

ORIGIN

Query Match 77.2%; Score 833.8; DB 10; Length 1488;
Best Local Similarity 86.9%; Pred. No. 9.9e-129;
Matches 932; Conservative 0; Mismatches 132; Indels 9; Gaps 1;

```
Qy          8 AGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGTGCTGCTG 67
              ||||| ||||| |||| |||||||||||||||||||||||||||||||||||||||||||
Db        125 AGCCACTCTTGACAGAACATCAACACCAGCCTCAAACTAGCAACTGTACAGGTGCTGCTG 184

Qy         68 CTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTGCCCGAGG 127
              |||||| |||||||  | | ||||||||||||||| |||||||||| ||||||||| |||
Db        185 TTGTCCATGAAGAGCATACCTCTGAGCGCCCCCCAAGCGCGGAGGAACGGGTGCCCAAGG 244

Qy        128 AGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGGCCGGAGG 187
              |||||||||||||||||||||||||| |  ||||  ||| ||||||| |||| |||||
Db        245 AGGACAGTAGGTGGCAATCGAGAGCGTCCTTGCAGTCCGGTAGCCGTCCAGGGCAGGAGG 304

Qy        188 GGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCAGAATCTA 247
              | |||||| ||||| |  ||| |  || || ||||||||  | ||||||||||||  |
Db        305 GAGAAGGGGGCCTGAAGCACCAGCTGCCCCCATTGCAGACGAATGCCTGTCCAGAATTGA 364

Qy        248 GCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGCGGCGACT 307
              ||| ||||    |||||||||||||||||||||||| ||||  ||| ||||||| | ||
Db        365 GCTCCCTGGAGAAGGGCGAGAAGGGCCAGAATGGGGAGGACTTATCCACTGGCGGTGCCT 424

Qy        308 TCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAGCCTTGTC 367
              ||||  || |||  | ||           |||   | | |||   ||||  |||||  |||
Db        425 CCCCGTCGGCGGAGGGAGA---------GCCGATGTCAGAGTCCCTCGTGCAGCCAGGTC 475

Qy        368 ATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAGTGGGGAC 427
              |||||||  |||||||   |||   || |||||||||| || |||||  |||   |||||||
Db        476 ATGACTCGGAGGCCACCAAGCAGGAGGCTCCTGCCGCCGGAGGCGAGGAGCCATGGGGAC 535

Qy        428 AGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAGAAGCGGC 487
              |||| |||||||||||||| ||||||||||   ||||||||| || ||||||||||||||
Db        536 AGCAACAGAGACAGCTGGGCAAGAAAAAACATCGGAGACGCCCCTCAAAGAAGAAGCGGC 595

Qy        488 ATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAGAAACAGA 547
              |||||||  || ||||||||||||  ||||| ||||||||||||||||||||||| ||||
Db        596 ATTGGAAGCCCTACTACAAGCTGACTTGGGAGGAGAAGAAAAAGTTCGACGAGAAGCAGA 655

Qy        548 GCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTCGCGCCCT 607
              |||| ||||||||  || | ||||||||||||||||||||||||||||||||| || ||||||
Db        656 GCCTGCGAGCTTCGCGGGTTCGAGCCGAGATGTTCGCCAAGGGCCAGCCAGTTGCGCCCT 715

Qy        608 ATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTCAAAACCG 667
              |||||||||||||||||||||||||||| ||||| |||||||||| ||||||||||||
Db        716 ATAACACCACGCAGTTCCTCATGGATGACCACGATCAGGAGGAGCCTGATCTCAAAACCG 775

Qy        668 GCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGACTTCATGG 727
              |||| ||| |||||||||| ||||||||||||||||||||||||||| || ||  |||
Db        776 GCCTTTACCCCAAGCGGGCAGCCGCCAAATCCGACGACACCAGCGATGAGGATTTTGTGG 835

Qy        728 AAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGCGAGTTTC 787
              ||||||  |||||||| || || |||||||||| ||||||||||||||||||||||||||
Db        836 AAGAAGCTGGTGAGGAAGACGGAGGCAGCGATGGCATGGGAGGGGACGGCAGCGAGTTTC 895

Qy        788 TGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAGAACATGA 847
              |||||||||||||||||||||||||||||||||||| | ||||||||||||||||||
Db        896 TGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACGCCGAGAGCCTGCAGAACATGA 955

Qy        848 GCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGCATGGAGG 907
              |||||||||||||||||| |||||||||| |||||||||||||||| |||| |||||
Db        956 GCAAGCAGGAGCTCATCAAAGAGTACCTGGAGCTGGAGAAGTGCCTCTCCCGCAAGGAGG 1015

Qy        908 ACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCGCGTGTGC 967
              |||| || |||||||||||||||| |||||||||||||| |||| |||||||||| ||||
Db       1016 ACGAAAATAACCGGCTGCGGCTGGAAAGCAAGCGGCTGGGAGGCGTCGACGCGCGAGTGC 1075

Qy        968 GGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTGACCGAGA 1027
              ||||||| ||||| |||||||||||||||||| ||||||||||||||||||||||||||||
Db       1076 GGGAGCTCGAGCTAGAGCTGGACCGGCTGCGCGCTGAGAACCTCCAGCTGCTGACCGAGA 1135
```

```
Qy       1028 ACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080
              ||||||||||||||||||||||||||||||| |||||||||| || ||||||
Db       1136 ACGAACTGCACCGGCAGCAGGAGCGAGCGCCTCTTTCCAAGTTCGGCGACTAG 1188
```

RESULT 13
AC120950/c
LOCUS AC120950 213625 bp DNA linear HTG 19-NOV-2002
DEFINITION Rattus norvegicus clone CH230-456H21, WORKING DRAFT SEQUENCE, 5 unordered pieces.
ACCESSION AC120950
VERSION AC120950.4 GI:25091749
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 213625)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 213625)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (14-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 213625)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 19, 2002 this sequence version replaced gi:23907976.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

```
-------------- Genome Center
    Center: Baylor College of Medicine
    Center code: BCM
    Web site: http://www.hgsc.bcm.tmc.edu/
    Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
    Center project name: GWRQ
    Center clone name: CH230-456H21
-------------- Summary Statistics
    Assembly program: Phrap; version 0.990329
    Consensus quality: 196084 bases at least Q40
    Consensus quality: 197843 bases at least Q30
    Consensus quality: 198926 bases at least Q20
    Estimated insert size: 201713; sum-of-contigs estimation
    Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
--------------
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1    58094: contig of 58094 bp in length
*    58095    58194: gap of unknown length
*    58195   139497: contig of 81303 bp in length
*   139498   139597: gap of unknown length
*   139598   208746: contig of 69149 bp in length
*   208747   208846: gap of unknown length
*   208847   210051: contig of 1205 bp in length
*   210052   210151: gap of unknown length
*   210152   213625: contig of 3474 bp in length.
```

FEATURES Location/Qualifiers
source 1. .213625
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-456H21"
misc_feature 1. .1190
/note="wgs_end_extension
clone_end:Sp6"
misc_feature 47586. .47846
/note="clone_boundary
clone_end:Sp6
site:
end_sequence:BZ240544"
BASE COUNT 51942 a 49325 c 49186 g 48940 t 14232 others

ORIGIN

Query Match 75.7%; Score 818; DB 2; Length 213625;
Best Local Similarity 86.2%; Pred. No. 2.1e-126;
Matches 931; Conservative 0; Mismatches 140; Indels 9; Gaps 2;

```
Qy          1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60
              ||||| |||||| |||||||||| |||||||||||||||||||||||||||||||||||
Db     119733 ATGGCTGAGCCACTCTTGTCAGAACATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 119674

Qy         61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120
              ||||||| |||||| || |||| |||| |||||||||||||| |||||||||| ||||||
Db     119673 GCTGCTGTTGTCCATGAGGAGCAGAACTCTGAGCGCCCCCCAAGCGCGGAGGAACGGGTG 119614

Qy        121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180
              ||  |||||||||||||||||||||||||||||| |    |||| |||| ||||||| |||
Db     119613 CCTAAGGAGGACAGTAGGTGGCAATCGAGAGCGTCCTTGCAGTCGGGTAGCCGTCCAGGG 119554

Qy        181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240
              | |||||| || ||| ||||| |    ||| |  || || ||||||||  | |||||||||
Db     119553 CAGGAGGGAGACGGGGGCCTGAAGCACCAGCTGCCCCCATTGCAGACGAATGCCTGTCCA 119494

Qy        241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300
              ||||  |||||| ||  | |||||||||||||||||||||||| ||||  ||| |||||
Db     119493 GAATTGAGCTGCTTGGAAAAGGGCGAGAAGGGCCAGAATGGGGAGGACTTATCCACTGGC 119434

Qy        301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360
              || | ||   ||||  |||| || | ||| |||| | | ||| ||      ||||  ||
Db     119433 GGTGCCT---CGCCTTCGGCGGAGGGGGAGCCGATGTCAGAGTCC------CTCGTGCAA 119383

Qy        361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 420
              ||  |||||||||| ||||| | |||| ||| |||||||| ||| || |   || |||
Db     119382 CCAGGTCATGACTCGGAGGCTACTAAGCTGGAGGCTCCTGTCGCTGGAGTAGAGGAGCCA 119323

Qy        421 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAG 480
              |||||||||||| |||||||||||||| |||||||||    |||||||| |||||||||
Db     119322 TGGGGACAGCAACAGAGACAGCTGGGCAAGAAAAAACACCGGAGACGCCCCTCCAAGAAG 119263

Qy        481 AAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAG 540
              ||||| |||||||| || || |||||||||  ||||| ||||||||||||||||||||||
Db     119262 AAGCGACATTGGAAGCCCTATTACAAGCTGACTTGGGAGGAGAAGAAAAAGTTCGACGAG 119203

Qy        541 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 600
              || || ||||| |||||||  |  | |||||||||||||||||||||||||||||||||||
Db     119202 AAGCAAAGCCTGCGAGCTTCGCGAGTTCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTT 119143

Qy        601 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 660
              |||||||||||||||||||||||||||||||||| || || |||||||||| ||||||
Db     119142 GCGCCCTATAACACCACGCAGTTCCTCATGGATGACCATGATCAGGAGGAGCCTGATCTC 119083

Qy        661 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 720
              |||||||||| ||| | |||||||| |||||||||||||||||||||||||||||| ||
Db     119082 AAAACCGGCCTTTACCCTAAGCGGGCAGCCGCCAAATCCGACGACACCAGCGATGAGGAT 119023

Qy        721 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 780
              ||| |||||||||  |||||||| || || |||||||||| ||||||||||||||||||||
Db     119022 TTCGTGGAAGAAGCTGGTGAGGAAGACGGAGGCAGCGATGGCATGGGAGGGGACGGCAGC 118963

Qy        781 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 840
              |||||||||||||||||| || ||||| |||||||||||||| | ||||||||||||
Db     118962 GAGTTTCTGCAGCGGGACTTTTCCGAGACTTACGAGCGGTACCACGCCGAGAGCCTGCAG 118903

Qy        841 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 900
              |||||||||||||||||||||||| |||||||||| ||||||||||||||||| |||
Db     118902 AACATGAGCAAGCAGGAGCTCATCAAAGAGTACCTGGAGCTGGAGAAGTGCCTCTCCCGC 118843

Qy        901 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 960
              | |||||||||| || |||||||||| |||| |||||||||| |||||||||  ||| |||
Db     118842 AAGGAGGACGAAAATAACCGGCTGCGGTTGGAAAGCAAGCGGTTGGGTGGCGTCGATGCG 118783

Qy        961 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1020
              || |||||||||||||||| |||||||||| |||||||||||||||| ||||||||||
Db     118782 CGAGTGCGGGAGCTGGAGCTAGAGCTGGACCGACTGCGCGCCGAGAACCGCCAGCTGCTG 118723

Qy       1021 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080
              ||||| |||||||||||||||||||||||||||||| | ||||||||| || ||||||
Db     118722 ACCGAAAACGAACTGCACCGGCAGCAGGAGCGAGCGCCTCCTTCCAAGTTCGGCGACTAG 118663
```

RESULT 14
AC136172/c
LOCUS AC136172 111803 bp DNA linear HTG 30-OCT-2002
DEFINITION Rattus norvegicus clone CH230-226J2, *** SEQUENCING IN PROGRESS ***, 53 unordered pieces.
ACCESSION AC136172
VERSION AC136172.1 GI:24421642
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 111803)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., Mcleod,M., Mcneill,T., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 111803)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission

```
  JOURNAL   Submitted (30-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
  COMMENT   -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: KCVI
                Center clone name: CH230-226J2
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 67362 bases at least Q40
                Consensus quality: 72560 bases at least Q30
                Consensus quality: 77367 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 53 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1244: contig of 1244 bp in length
            *     1245     1344: gap of unknown length
            *     1345     2653: contig of 1309 bp in length
            *     2654     2753: gap of unknown length
            *     2754     4274: contig of 1521 bp in length
            *     4275     4374: gap of unknown length
            *     4375     5978: contig of 1604 bp in length
            *     5979     6078: gap of unknown length
            *     6079     7135: contig of 1057 bp in length
            *     7136     7235: gap of unknown length
            *     7236     8520: contig of 1285 bp in length
            *     8521     8620: gap of unknown length
            *     8621    10172: contig of 1552 bp in length
            *    10173    10272: gap of unknown length
            *    10273    11379: contig of 1107 bp in length
            *    11380    11479: gap of unknown length
            *    11480    12576: contig of 1097 bp in length
            *    12577    12676: gap of unknown length
            *    12677    13863: contig of 1187 bp in length
            *    13864    13963: gap of unknown length
            *    13964    14987: contig of 1024 bp in length
            *    14988    15087: gap of unknown length
            *    15088    16573: contig of 1486 bp in length
            *    16574    16673: gap of unknown length
            *    16674    17896: contig of 1223 bp in length
            *    17897    17996: gap of unknown length
            *    17997    19475: contig of 1479 bp in length
            *    19476    19575: gap of unknown length
            *    19576    20884: contig of 1309 bp in length
            *    20885    20984: gap of unknown length
            *    20985    22487: contig of 1503 bp in length
            *    22488    22587: gap of unknown length
            *    22588    24747: contig of 2160 bp in length
            *    24748    24847: gap of unknown length
            *    24848    26172: contig of 1325 bp in length
            *    26173    26272: gap of unknown length
            *    26273    27802: contig of 1530 bp in length
            *    27803    27902: gap of unknown length
            *    27903    29918: contig of 2016 bp in length
            *    29919    30018: gap of unknown length
            *    30019    31165: contig of 1147 bp in length
            *    31166    31265: gap of unknown length
            *    31266    32793: contig of 1528 bp in length
            *    32794    32893: gap of unknown length
            *    32894    34401: contig of 1508 bp in length
            *    34402    34501: gap of unknown length
            *    34502    35644: contig of 1143 bp in length
            *    35645    35744: gap of unknown length
            *    35745    36968: contig of 1224 bp in length
            *    36969    37068: gap of unknown length
            *    37069    38349: contig of 1281 bp in length
            *    38350    38449: gap of unknown length
            *    38450    41409: contig of 2960 bp in length
            *    41410    41509: gap of unknown length
            *    41510    43066: contig of 1557 bp in length
            *    43067    43166: gap of unknown length
            *    43167    44827: contig of 1661 bp in length
            *    44828    44927: gap of unknown length
            *    44928    46592: contig of 1665 bp in length
            *    46593    46692: gap of unknown length
            *    46693    49255: contig of 2563 bp in length
            *    49256    49355: gap of unknown length
            *    49356    51247: contig of 1892 bp in length
            *    51248    51347: gap of unknown length
            *    51348    53114: contig of 1767 bp in length
            *    53115    53214: gap of unknown length
            *    53215    55332: contig of 2118 bp in length
            *    55333    55432: gap of unknown length
            *    55433    57427: contig of 1995 bp in length
            *    57428    57527: gap of unknown length
            *    57528    59606: contig of 2079 bp in length
            *    59607    59706: gap of unknown length
            *    59707    61138: contig of 1432 bp in length
            *    61139    61238: gap of unknown length
            *    61239    63710: contig of 2472 bp in length
            *    63711    63810: gap of unknown length
            *    63811    66371: contig of 2561 bp in length
            *    66372    66471: gap of unknown length
            *    66472    70029: contig of 3558 bp in length
            *    70030    70129: gap of unknown length
            *    70130    72934: contig of 2805 bp in length
            *    72935    73034: gap of unknown length
            *    73035    76024: contig of 2990 bp in length
            *    76025    76124: gap of unknown length
            *    76125    79346: contig of 3222 bp in length
            *    79347    79446: gap of unknown length
            *    79447    82950: contig of 3504 bp in length
            *    82951    83050: gap of unknown length
            *    83051    86116: contig of 3066 bp in length
            *    86117    86216: gap of unknown length
            *    86217    89119: contig of 2903 bp in length
            *    89120    89219: gap of unknown length
            *    89220    92494: contig of 3275 bp in length
            *    92495    92594: gap of unknown length
            *    92595    94419: contig of 1825 bp in length
            *    94420    94519: gap of unknown length
            *    94520    98048: contig of 3529 bp in length
            *    98049    98148: gap of unknown length
            *    98149   101599: contig of 3451 bp in length
            *   101600   101699: gap of unknown length
            *   101700   105918: contig of 4219 bp in length
            *   105919   106018: gap of unknown length
            *   106019   108449: contig of 2431 bp in length
            *   108450   108549: gap of unknown length
            *   108550   111803: contig of 3254 bp in length.
  FEATURES             Location/Qualifiers

    Query Match             43.6%;  Score 471.2;  DB 2;  Length 111803;
    Best Local Similarity   90.3%;  Pred. No. 7.8e-69;
    Matches  514;  Conservative    0;  Mismatches   54;  Indels    1;  Gaps    1;

  Qy        512 ACTGGGAAGAGAAGAAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGAG 571
                |||  ||| ||||||||||||||||||||||||| || ||||| ||||||||  |  | ||||
  Db      14984 ACTTGGAGGAGAAGAAAAAGTTCGACGAGAAGCAAAGCCTGCGAGCTTCGCGAGTTCGAG 14925

  Qy        572 CCGAGATGTTCGCCAAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTCATGG 631
                ||||||||  |||||||||||||||||||| ||||||||||||||||||||||||| ||||||
```

```
Db      14924 CCGAGATG-TCGCCAAGGGCCAGCCGGTTGCGCCCTATAACACCACGCAGTTCNTCATGG 14866

Qy        632 ATGATCACGACCAGGAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCCG 691
              |||| || || |||||||||| |||||||||||||||| ||| | |||||||| ||||
Db      14865 ATGACCATGATCAGGAGGAGCCTGATCTCAAAACCGGCCTTTACCCTAAGCGGGCAGCCG 14806

Qy        692 CCAAATCCGACGACACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGG 751
              |||||||||||||||||||||||| || ||| ||||||||  |||||||| || || |
Db      14805 CCAAATCCGACGACACCAGCGATGAGGATTTCGTGGAAGAAGCTGGTGAGGAAGACGGAG 14746

Qy        752 GCAGCGATGGGATGGGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGT 811
              |||||||||| |||||||||||||||||||||||||||||||||||| || ||||| |
Db      14745 GCAGCGATGGCATGGGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTTTCCGAGACTT 14686

Qy        812 ACGAGCGGTACCACACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGT 871
              |||||||||||||| | |||||||||||||||||||||||||||||||||||||| ||||
Db      14685 ACGAGCGGTACCACGCCGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAAGAGT 14626

Qy        872 ACCTGGAACTGGAGAAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGG 931
              ||||||| |||||||||||||||| |||| ||||||||| || |||||||||||| |||
Db      14625 ACCTGGAGCTGGAGAAGTGCCTCTCCCGCAAGGAGGACGAAAATAACCGGCTGCGGTTGG 14566

Qy        932 AGAGCAAGCGGCTGGGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACC 991
              | ||||||||| |||||||| ||| ||||| |||||||||||||||| ||||||||||
Db      14565 AAAGCAAGCGGTTGGGTGGCGTCGATGCGCGAGTGCGGGAGCTGGAGCTAGAGCTGGACC 14506

Qy        992 GGCTGCGCGCCGAGAACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGC 1051
              | |||||||||||||||| |||||||||||||| |||||||||||||||||||||||||
Db      14505 GACTGCGCGCCGAGAACCGCCAGCTGCTGACCGAAAACGAACTGCACCGGCAGCAGGAGC 14446

Qy       1052 GAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080
              ||||||| | ||||||||| || ||||||
Db      14445 GAGCGCCTCCTTCCAAGTTCGGCGACTAG 14417
```

```
RESULT 15
BD058159
LOCUS       BD058159                523 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Secreted expressed sequence tags (sESTs).
ACCESSION   BD058159
VERSION     BD058159.1  GI:22603765
KEYWORDS    JP 2001519666-A/14.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 523)
  AUTHORS   Jacobs,K., Mccoy,J.M., Lavallie,E.R., Racie,L.A., Merberg,D.,
            Treacy,M., Spaulding,V. and Agostino,M.J.
  TITLE     Secreted expressed sequence tags (sESTs)
  JOURNAL   Patent: JP 2001519666-A 14 23-OCT-2001;
            GENETICS INSTITUTE INC
COMMENT     PN   JP 2001519666-A/14
            PD   23-OCT-2001
            PF   10-APR-1998 JP 1998543068
            PR   10-APR-1997 US    08/835913
            PI   KENNETH JACOBS,JOHN M MCCOY,EDWARD R LAVALLIE,LISA A RACIE, PI
              DAVID MERBERG,
            PI   MAURICE TREACY,VIKKI SPAULDING,MICHAEL J AGOSTINO PC
            C12N15/12,C12N5/10,C07K14/47,C12Q1/68,A61K38/17 CC   Strandedness:
            Double;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers.
FEATURES            Location/Qualifiers
     source         1. .523
                    /organism="Zea mays"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:4577"
BASE COUNT      112 a    149 c    153 g    108 t      1 others
ORIGIN
```

Query Match 36.6%; Score 395.8; DB 6; Length 523;

Best Local Similarity 96.6%; Pred. No. 5.3e-56;
Matches 403; Conservative 1; Mismatches 13; Indels 0; Gaps 0;

```
Qy          1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        104 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 163

Qy         61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        164 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 223

Qy        121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        224 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 283

Qy        181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        284 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 343

Qy        241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        344 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 403

Qy        301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360
              ||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        404 GGCGACTTYCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 463

Qy        361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAG 417
              |||||||||||||||||||||||||| ||||| |  | ||| |  || |   |||
Db        464 CCTTGTCATGACTCCGAGGCCAGTAAGATGGGGTCGACGGCCTCTTTGGCCCTCGAG 520
```

Search completed: February 5, 2004, 11:18:36
Job time : 4192 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: February 5, 2004, 13:10:24 ; Search time 110 Seconds
(without alignments)
842.190 Million cell updates/sec

Title: US-09-972-758A-2
Perfect score: 1910
Sequence: 1 MAEPFLSEYQHQPQTSNCTG.....LTENELHRQQERAPLSKFGD 359

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1910	100.0	359	4	O94992	O94992 homo sapien
2	1587.5	83.1	356	11	Q8R409	Q8r409 mus musculu
3	1199	62.8	246	11	Q8VDF7	Q8vdf7 mus musculu
4	545.5	28.6	286	4	Q96MH2	Q96mh2 homo sapien
5	483.5	25.3	280	11	Q9D4C7	Q9d4c7 mus musculu
6	213	11.2	360	5	Q8INF6	Q8inf6 drosophila
7	212	11.1	360	5	Q8IGW8	Q8igw8 drosophila
8	209	10.9	349	5	Q9VFH2	Q9vfh2 drosophila
9	150	7.9	1142	5	Q8T6B4	Q8t6b4 dictyosteli
10	148.5	7.8	992	4	Q9NTH6	Q9nth6 homo sapien
11	147	7.7	1455	4	Q9UPV0	Q9upv0 homo sapien
12	142.5	7.5	560	6	Q95LS7	Q95ls7 macaca fasc
13	141.5	7.4	1132	5	Q9W475	Q9w475 drosophila
14	140.5	7.4	443	5	O76153	O76153 periplaneta
15	140	7.3	609	17	Q8TXA4	Q8txa4 methanopyru
16	140	7.3	714	10	Q94DE3	Q94de3 oryza sativ
17	139	7.3	1781	4	Q9UKX0	Q9ukx0 homo sapien
18	139	7.3	1890	4	Q9UKW3	Q9ukw3 homo sapien
19	139	7.3	2073	4	Q9UKW2	Q9ukw2 homo sapien
20	138.5	7.3	695	11	Q8CI48	Q8ci48 mus musculu
21	138.5	7.3	944	11	Q922S3	Q922s3 mus musculu
22	138.5	7.3	1665	11	Q8VII1	Q8vii1 mus musculu
23	138	7.2	2072	4	Q8WYB5	Q8wyb5 homo sapien
24	137.5	7.2	483	6	Q8MJG1	Q8mjg1 bos taurus
25	137.5	7.2	540	11	Q9EQ53	Q9eq53 mus musculu
26	136	7.1	399	10	Q94IK2	Q94ik2 solanum tub
27	135.5	7.1	530	4	Q9UER6	Q9uer6 homo sapien
28	135.5	7.1	1402	3	Q8NIV6	Q8niv6 neurospora
29	135	7.1	540	4	Q96B87	Q96b87 homo sapien
30	135	7.1	554	4	Q9HBQ2	Q9hbq2 homo sapien
31	135	7.1	555	4	Q8N2I7	Q8n2i7 homo sapien
32	135	7.1	602	3	Q9HEL8	Q9hel8 neurospora
33	135	7.1	813	4	Q9ULN3	Q9uln3 homo sapien
34	133	7.0	462	11	Q9R1B7	Q9r1b7 rattus norv
35	133	7.0	538	15	Q9WJP4	Q9wjp4 moloney mur
36	133	7.0	1144	5	Q9NJH7	Q9njh7 drosophila
37	133	7.0	1737	15	O92808	O92808 moloney mur
38	132.5	6.9	530	4	O75475	O75475 homo sapien
39	132.5	6.9	911	5	Q8T852	Q8t852 dictyosteli
40	132	6.9	528	4	Q9H6U3	Q9h6u3 homo sapien
41	132	6.9	528	4	Q9H3P7	Q9h3p7 homo sapien
42	132	6.9	528	4	Q8IZC5	Q8izc5 homo sapien
43	132	6.9	2400	4	Q8IWP2	Q8iwp2 homo sapien
44	132	6.9	2416	4	Q8IWP1	Q8iwp1 homo sapien
45	132	6.9	2432	4	Q8IWP0	Q8iwp0 homo sapien

ALIGNMENTS

RESULT 1
O94992
ID O94992 PRELIMINARY; PRT; 359 AA.
AC O94992;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HIS1 protein (HMBA-inducible).
GN HIS1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kusuhara M., Nagasaki K., Kimura K., Ishikawa S., Maass N., Ogawa S.,
RA Yamaguchi K.;
RT "Cloning of HMBA-inducible transcript, HIS1, in human vascular smooth
RT muscle cells.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB021179; BAA36166.1; -.
DR EMBL; BC006460; AAH06460.1; -.
SQ SEQUENCE 359 AA; 40623 MW; B12845C4E2595FF0 CRC64;

Query Match 100.0%; Score 1910; DB 4; Length 359;
Best Local Similarity 100.0%; Pred. No. 9.5e-132;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAEPFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPG 60
Db 1 MAEPFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPG 60

Qy 61 PEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELLAQ 120

```
Db          61 PEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELLAQ 120

Qy         121 PCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDE 180

Db         121 PCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDE 180

Qy         181 KQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDD 240

Db         181 KQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDD 240

Qy         241 FMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCLSR 300

Db         241 FMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCLSR 300

Qy         301 MEDENNRLRLESKRLGGDDARVRELELELDRLRAENLQLLTENELHRQQERAPLSKFGD 359

Db         301 MEDENNRLRLESKRLGGDDARVRELELELDRLRAENLQLLTENELHRQQERAPLSKFGD 359


RESULT 2
Q8R409
ID   Q8R409      PRELIMINARY;      PRT;   356 AA.
AC   Q8R409;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Cardiac lineage protein 1.
GN   CLP1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SvJ;
RA   Huang F., Wagner M., Siddiqui M.;
RT   "Structure, expression, and functional characterization of the mouse
RT   CLP-1 gene.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY090614; AAM09026.1; -.
DR   MGD; MGI:2385923; Clp1.
SQ   SEQUENCE   356 AA;  40243 MW;  242DE7EE66BA293B CRC64;

  Query Match             83.1%;  Score 1587.5;  DB 11;  Length 356;
  Best Local Similarity   85.8%;  Pred. No. 3.5e-108;
  Matches  308;  Conservative   11;  Mismatches   37;  Indels    3;  Gaps    2;

Qy           1 MAEPFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPG 60

Db           1 MAEPLLTEHQHQPQTSNCTGAAVVHEEHTSERPPSAEERVPKEDSRWQSRASLQSGSRPG 60

Qy          61 PEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELLAQ 120

Db          61 QEGEGGLKHQLPPLQTNACPELSSLEKGEKGQNGEDLSTGG-ASPSAEGEPMSES--LVQ 117

Qy         121 PCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDE 180

Db         118 PGHDSEATKQEAPAAGGEEPWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDE 177

Qy         181 KQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDD 240

Db         178 KQSLRASRVRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYPKRAAAKSDDTSDED 237

Qy         241 FMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCLSR 300

Db         238 FVEEAGEEDGGSDGMGGDGSEFLQRDFSETYERYHAESLQNMSKQELIKEYLELEKCLSR 297

Qy         301 MEDENNRLRLESKRLGGDDARVRELELELDRLRAENLQLLTENELHRQQERAPLSKFGD 359

Db         298 KEDENNRLRLESKRLGGVDARVRELELELDRLRAENLQLLTENELHRQQERAPLSKFGD 356


RESULT 3
Q8VDF7
ID   Q8VDF7      PRELIMINARY;      PRT;   246 AA.
AC   Q8VDF7;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to HMBA-inducible.
GN   CLP1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC022111; AAH22111.1; -.
DR   MGD; MGI:2385923; Clp1.
SQ   SEQUENCE   246 AA;  28572 MW;  6CF8227F599B77EF CRC64;

  Query Match             62.8%;  Score 1199;  DB 11;  Length 246;
  Best Local Similarity   93.5%;  Pred. No. 5.5e-80;
  Matches  229;  Conservative    5;  Mismatches   11;  Indels    0;  Gaps    0;

Qy         115 AELLAQPCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEE 174

Db           2 SESLVQPGHDSEATKQEAPAAGGEEPWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEE 61

Qy         175 KKKFDEKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSD 234

Db          62 KKKFDEKQSLRASRVRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYPKRAAAKSD 121

Qy         235 DTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLEL 294

Db         122 DTSDEDFVEEAGEEDGGSDGMGGDGSEFLQRDFSETYERYHAESLQNMSKQELIKEYLEL 181

Qy         295 EKCLSRMEDENNRLRLESKRLGGDDARVRELELELDRLRAENLQLLTENELHRQQERAPL 354

Db         182 EKCLSRKEDENNRLRLESKRLGGVDARVRELELELDRLRAENLQLLTENELHRQQERAPL 241

Qy         355 SKFGD 359

Db         242 SKFGD 246


RESULT 4
Q96MH2
ID   Q96MH2      PRELIMINARY;      PRT;   286 AA.
AC   Q96MH2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein FLJ32384 (Similar to putative).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skeletal muscle;
RA   Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA   Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA   Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,
RA   Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Eye;
RA   Strausberg R.;
```

```
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK056946; BAB71319.1; -.
DR   EMBL; BC025970; AAH25970.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   286 AA;  32418 MW;  58576D72096A8A6F CRC64;

  Query Match           28.6%;  Score 545.5;  DB 4;  Length 286;
  Best Local Similarity   43.8%;  Pred. No. 3.7e-32;
  Matches  128;  Conservative   36;  Mismatches   93;  Indels   35;  Gaps    6;

Qy         73 PLQTQACPESSCLREGEKGQNGDDS--------SAGGDFPPPAEVEPTPEAELLAQPC-- 122
              | ||    ||    |  |       |         :||  |    :|   | | ||
Db          5 PNQTACNAESPVALEEAKTSGAPGSPQTPPERHDSGGSLPLTPRMESHSEDEDLAGAVGG 64

Qy        123 --HDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDE 180
                 :| : :  :|     |         | :|||||||||:||||:|| :|:|  ||:: ||
Db         65 LGWNSRSPRTQSPGGCSAE------AVLARKKHRRRPSKRKRHWRPYLELSWAEKQQRDE 118

Qy        181 KQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEP--DLKTGLYSKRAAAKSDDTSD 238
              :|| ||||:| ||||||||||||||||||:| | |||  |:  |:    :: :|
Db        119 RQSQRASRVREEMFAKGQPVAPYNTTQFLMNDRDPEEPNLDVPHGISHPGSSGES----- 173

Qy        239 DDFMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCL 298
                    | | ||| |    || ::|||||||:|||||  |||||:::|||||| |
Db        174 ---------EAGDSDGRGRAHGEFQRKDFSETYERFHTESLQGRSKQELVRDYLELEKRL 224

Qy        299 SRMEDENNRLRLESKRLGGDDAR-VRELELELDRLRAENLQLLTENELHRQQ 349
              |: |:|  ||:     |    | | ||  |: ||| || :|  ||::  ::
Db        225 SQAEEETRRLQQLQACTGQQSCRQVEELAAEVQRLRTENQRLRQENQMWNRE 276


RESULT 5
Q9D4C7
ID   Q9D4C7      PRELIMINARY;      PRT;   280 AA.
AC   Q9D4C7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   4933402L21Rik protein.
GN   4933402L21RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK016624; BAB30344.1; -.
DR   MGD; MGI:1918309; 4933402L21Rik.
SQ   SEQUENCE   280 AA;  32388 MW;  6BA925FE41A63234 CRC64;

  Query Match           25.3%;  Score 483.5;  DB 11;  Length 280;
  Best Local Similarity   44.4%;  Pred. No. 1.2e-27;
  Matches  112;  Conservative   29;  Mismatches   66;  Indels   45;  Gaps    7;

Qy        122 CHDSEASK----------------LGA--PAAGGEEEWGQQQRQLGKKKHRRRPSKKKRH 163
              ||   |                 ||   | | |    |    |:   ::       : :||
Db         12 CHQDRRSNQRMKGLSPLAMAQVVTLGVLEPRAQGAVRW----RRCWPERSTVGAIEAQRH 67

Qy        164 WKPYYKLTWEEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTG 223
              |:|| :|:| ||:: ||:|| ||||:| ||||||||:||||||||||:| | |||:|
Db         68 WRPYLELSWAEKQQRDERQSQRASRVREEMFAKGQPLAPYNTTQFLMNDRDLEEPNL--- 124

Qy        224 LYSKRAAAKSDDTSDDDFMEEGGE-EDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNM 282
                          |           ||| | | ||| |    || |||||| ||||||||||
Db        125 ----------DVLHGPSHSGSGGENEAGDSDGQGRAHGEFQQRDFSEAYERYHTESLQGR 174

Qy        283 SKQELIKEYLELEKCLSRMEDENNRLRLESKRLGGDDAR-----VRELELELDRLRAENL 337
              |||||:::||:||: ||: | |  |||    :| |  :|     | ||  |::||| ||
Db        175 SKQELVRDYLDLERRLSQAEQETRRLR----QLQGCSSRQPCQQVEELAAEVERLRTENQ 230

Qy        338 QLLTENELHRQQ 349
              :|  |||:  ::
Db        231 RLRQENEMWNRE 242


RESULT 6
Q8INF6
ID   Q8INF6      PRELIMINARY;      PRT;   360 AA.
AC   Q8INF6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CG3508-PB.
GN   CG3508.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
```

```
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003705; AAN13614.1; -.
SQ   SEQUENCE   360 AA;  39211 MW;  41F350CA35965233 CRC64;

  Query Match           11.2%;  Score 213;  DB 5;  Length 360;
  Best Local Similarity  27.8%;  Pred. No. 9.6e-08;
  Matches   69;  Conservative   38;  Mismatches   73;  Indels   68;  Gaps   10;

Qy        125 SEASKLGAPAAGG--EEEWGQQQRQLG----------------------KKKHRR----- 155
              :|| | |      |  |:| | ||| |                        |:|||
Db          2 AEAVKNGKNRHSGSAEKESGSQQRPLDSGGGGGASGGGGVAVGGGSGMPKRKHRRGKKSK 61

Qy        156 -RPSKKKRHWKPYYKLTWE----EKKKFDEKQSLRASRIRAEMFAKGQPVAPYNTTQFLM 210
               :| | | |: | :||         : :::|: |   :|:        : |||| :|||
Db         62 MQPKKTKNHY-PQWKLDMSTGAGATLEGNQRQNSRTKLVRSRSL-----LVPYNTNRFLM 115

Qy        211 DDHDQEEPDLKTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSET 270
              ::|  |                 ||: |: |   | : :        |   || ::|:
Db        116 EEHMSE-------------LHKDDSDDNCF---GSQTE--------DQVLFLSKEFSDV 150

Qy        271 YERYHTESLQNMSKQELIKEYLELEKCLSRMEDENNRLRLESKRLGGDDARVRELELELD 330
              |||   | |: |||||||:| :::|   |: :: :        :|   | ::|:|  |
Db        151 YERARLERLETMSKQELIQECMQIEDRYSKAQNISKEF---GAKLRAQDDKIRQLSRENQ 207

Qy        331 RLRAENLQ 338
               ||   |:
Db        208 FLRTHLLR 215


RESULT 7
Q8IGW8
ID   Q8IGW8      PRELIMINARY;      PRT;   360 AA.
AC   Q8IGW8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   RE11603p.
GN   CG3508.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=y;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BT001551; AAN71306.1; -.
SQ   SEQUENCE   360 AA;  39182 MW;  5AEB437A3582A332 CRC64;

  Query Match           11.1%;  Score 212;  DB 5;  Length 360;
  Best Local Similarity  27.8%;  Pred. No. 1.1e-07;
  Matches   69;  Conservative   38;  Mismatches   73;  Indels   68;  Gaps   10;

Qy        125 SEASKLGAPAAGG--EEEWGQQQRQLG----------------------KKKHRR----- 155
              :|| | |      |  |:| | ||| |                        |:|||
Db          2 AEAVKNGKNRHSGSAEKESGSQQRPLDSGGGGGASGGGGVAVGGGSGMPKRKHRRGKKSK 61

Qy        156 -RPSKKKRHWKPYYKLTWE----EKKKFDEKQSLRASRIRAEMFAKGQPVAPYNTTQFLM 210
               :| | | |: | :||         : :::|: |   :|:        : |||| :|||
Db         62 MQPKKTKNHY-PQWKLDMSTGAGATLEGNQRQNSRTKLVRSRSL-----LVPYNTDRFLM 115

Qy        211 DDHDQEEPDLKTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSET 270
              ::|  |                 ||: |: |   | : :        |   || ::|:
Db        116 EEHMSE-------------LHKDDSDDNCF---GSQTE--------DQVLFLSKEFSDV 150

Qy        271 YERYHTESLQNMSKQELIKEYLELEKCLSRMEDENNRLRLESKRLGGDDARVRELELELD 330
              |||   | |: |||||||:| :::|   |: :: :        :|   | ::|:|  |
Db        151 YERARLERLETMSKQELIQECMQIEDRYSKAQNISKEF---GAKLRAQDDKIRQLSRENQ 207

Qy        331 RLRAENLQ 338
               ||   |:
Db        208 FLRTHLLR 215


RESULT 8
Q9VFH2
ID   Q9VFH2      PRELIMINARY;      PRT;   349 AA.
AC   Q9VFH2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   CG3508 protein (LD30520P).
GN   CG3508.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
```

```
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003705; AAF55085.1; -.
DR   EMBL; AY051786; AAK93210.1; -.
DR   FlyBase; FBgn0038251; CG3508.
SQ   SEQUENCE   349 AA;  38058 MW;  8597FC7AE08D0747 CRC64;

  Query Match             10.9%;  Score 209;  DB 5;  Length 349;
  Best Local Similarity   25.9%;  Pred. No. 1.8e-07;
  Matches   68;  Conservative   43;  Mismatches   84;  Indels   68;  Gaps   10;

Qy          85 LREGEKGQNGD-----DSSAGGDFPPPAEVEPTPEAELLAQPCHDSEASKLGAPAAGGEE 139
               : |  | ::|      ||  ||                          ||  |  | ||
Db           1 MAEAVKNESGSQQRPLDSGGGGG----------------------ASGGGGVAVGGGS 36

Qy         140 EWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWE----EKKKFDEKQSLRASRIRAEMFA 195
                ::: : |||  : :| | | |: | :||          : :::|: |   :|:
Db          37 GMPKRKHRRGKKS-KMQPKKTKNHY-PQWKLDMSTGAGATLEGNQRQNSRTKLVRSRSL- 93

Qy         196 KGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSDGM 255
                 : |||| :|||::|  |                  ||: |: |   | : :
Db          94 ----LVPYNTNRFLMEEHMSE--------------LHKDDSDDNCF---GSQTE------ 126

Qy         256 GGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCLSRMEDENNRLRLESKRL 315
                 |   || ::||: |||   | |: |||||||:| :::|   |: :: :        :|
Db         127 --DQVLFLSKEFSDVYERARLERLETMSKQELIQECMQIEDRYSKAQNISKEF---GAKL 181

Qy         316 GGDDARVRELELELDRLRAENLQ 338
                 | ::|:|  |   ||   |:
Db         182 RAQDDKIRQLSRENQFLRTHLLR 204


RESULT 9
Q8T6B4
ID   Q8T6B4      PRELIMINARY;      PRT;  1142 AA.
AC   Q8T6B4;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Non-transporter ABC protein AbcF4.
GN   ABCF4.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Ax4;
RA   Anjard C., Loomis W.F.;
RT   "Evolution of the ABC transporters of Dictyostelium.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF479256; AAL87694.1; -.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transporter.
DR   Pfam; PF00005; ABC_tran; 2.
DR   ProDom; PD000006; ABC_transporter; 2.
DR   SMART; SM00382; AAA; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW   ATP-binding.
SQ   SEQUENCE   1142 AA;  130224 MW;  33B1815AB09942DC CRC64;

  Query Match             7.9%;  Score 150;  DB 5;  Length 1142;
  Best Local Similarity   19.8%;  Pred. No. 0.015;
  Matches   79;  Conservative   73;  Mismatches  132;  Indels  114;  Gaps   18;

Qy          30 PERPPGAEERVPEEDSRWQSRAFPQ---LGGRPGPEGEGSLESQ--------PPPLQTQA 78
               |::  | :::  ::||  :    ||    ||:| |: :|  :           | |::
Db         132 PQKKGGKQQQ--QQDSDDEQEEIPQPVKKGGKPAPQKKGGKQQDSDDEEDEIPQPVKKGG 189

Qy          79 CPESSCLREGEKGQNGDDSSAGGDFPPPA----------------EVEPTPEAELLAQPC 122
                 |     ::| | |  :|     :   |                  | |   | | : ||
Db         190 KPAPQ--KKGGKQQESEDEDEEDEVQQPVKKGGKNDKKKGVKHVEEEEEEEEEEEIEQPV 247

Qy         123 HDSEASKLGAPAAGG---------EEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWE 173
                 :  |   | ||          ||:  ||  : | || ::: ||              |
Db         248 --KKGGKAPKPKKGGKGSKQESEDEEDDVQQPVKKGGKKDKKKGSKHV-----------E 294

Qy         174 EKKKFDEKQSLR------ASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGL--- 224
               |::: :|:: :        :::   :   ||: |           :: ::|| :::  :
Db         295 EEEEEEEEEEIEQPVKKGSNKKDQKKGGKGKHVEE--------EEEEEEEEEIEQPVKKG 346

Qy         225 -----YSKRAAAKSDDTSDDDFME------EGGEEDGGSDGMGGDGSEFLQRDFSETYER 273
                      |    |   |:|:  |       :||::|          ||: :: :  |  |
Db         347 SNKKDQKKGGKGKQQQESEDEEEEIQQPVKKGGKKDKKK------GSKHVEEEEEEEEE 400

Qy         274 YHTESLQNMSKQELIKEYLELEKCLSRMEDENNRLRLESKRLGGDDARVRELELE----- 328
                    |        : :|:  : :|  | :||  : | ::||: ||    | | | :
Db         401 EEIE--------QPVKKGGKKDK-KSSLEDSMSELSIKSKK-GGKGKHVEEEEEQEQEEE 450

Qy         329 ----------LDRLRAENLQLLTENELHRQQERAPLSK 356
                        |: : ::::   | |   ::|  | ||
Db         451 EEKPKSKSNKKDKKKGKHVE--EEEEEEEEEEKPKSK 486


RESULT 10
Q9NTH6
ID   Q9NTH6      PRELIMINARY;      PRT;   992 AA.
AC   Q9NTH6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein (Fragment).
GN   DKFZP434O2413.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Koehrer K., Beyer A., Mewes H.W., Gassenhuber J., Wiemann S.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL137265; CAB70664.1; -.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   992 AA;  112628 MW;  039DF5B1E97EF02F CRC64;

  Query Match             7.8%;  Score 148.5;  DB 4;  Length 992;
  Best Local Similarity   21.1%;  Pred. No. 0.016;
  Matches   94;  Conservative   66;  Mismatches  149;  Indels  137;  Gaps   17;

Qy         25 QEELNPERPPGAEERVP-----EEDSRWQSRAF------PQLGGRPGPEGE--------- 64
              |:||  ::  | |||:       || ::   |:        ||     ||||:
Db        230 QDELQSKQSKGLEERLSPPLPHEERAQSPPRSLATEEEPPQ-----GPEGQPEWKEAEEL 284

Qy         65 -----GSLESQPPPLQTQA-CPESSCLREGEKGQNGDDSSAGGDFPPPAE----VEPTP- 113
                  ||  |    : ||  | ::| :  |:     ::   | :     |    | |||
Db        285 GEDSAASLSLQLSLQREQAPSPPAACEKGKEQHSQAEELGPGQEEAEDPEEKVAVSPTPP 344

Qy        114 ------EAELLAQPCHDSEASKLGAPAAGGEEEWGQQQRQLGKKK--------------- 152
                  | :| |   |||: | |      :   | || | : |
Db        345 VSPEVRSTEPVAPPEQLSEAA-LKAMEEAVAQVLEQDQRHLLESKQEKMQQLREKLCQEE 403

Qy        153 -------HRRRPSKKKRHWKPYYKLTWEEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNT 205
                    |:::        :   |   ||: :  |::| | | :||:: :  |
Db        404 EEEILRLHQQKEQSLSSLRERLQKAIEEEEARMREEESQRLSWLRAQVQSSTQA------ 457

Qy        206 TQFLMDDHDQEEPDLKTGLYSKRAAAKSD--------DTSDDDFMEEGGEEDGGS----- 252
                  | ||   : :  |   |   :|        :  :   :|:  ||   |
Db        458 ------DEDQIRAEQEASLQKLREELESQQKAERASLEQKNRQMLEQLKEEIEASEKSEQ 511

Qy        253 -----------------DGMGGDGSEFLQRDFSETYERY-------HTE---SLQNMSK 284
                            :|   :    |::: |   ||         | |   |||   :
Db        512 AALNAAKEKALQQLREQLEGERKEAVATLEKEHSAELERLCSSLEAKHREVVSSLQKKIQ 571

Qy        285 QELIKEYLELEKCLSRMEDENNRLRLESKRLGGDDARVREL------------ELELDRL 332
                :   ||  :|:||| ::|   :|:  :|  : | :  :  |             |  ||::
Db        572 EAQQKEEAQLQKCLGQVE---HRVHQKSYHVAGYEHELSSLLREKRQEVEGEHERRLDKM 628

Qy        333 RAENLQLLT----ENELHRQQERAPL 354
                : |: |::     : |   :::|| |
Db        629 KEEHQQVMAKAREQYEAEERKQRAEL 654


RESULT 11
Q9UPV0
ID   Q9UPV0      PRELIMINARY;      PRT;  1455 AA.
AC   Q9UPV0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein KIAA1052.
GN   KIAA1052.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99397452; PubMed=10470851;
RA   Kikuno R., Nagase T., Ishikawa K., Hirosawa M., Miyajima N.,
RA   Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XIV.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:197-205(1999).
DR   EMBL; AB028975; BAA83004.1; -.
DR   InterPro; IPR001202; WW_Rsp5_WWP.
DR   Pfam; PF00397; WW; 1.
DR   SMART; SM00456; WW; 1.
DR   PROSITE; PS50020; WW_DOMAIN_2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   1455 AA;  163543 MW;  7F48093100C34819 CRC64;

  Query Match             7.7%;  Score 147;  DB 4;  Length 1455;
  Best Local Similarity   21.6%;  Pred. No. 0.033;
  Matches   97;  Conservative   65;  Mismatches  146;  Indels  142;  Gaps   19;

Qy         25 QEELNPERPPGAEER-------VPEEDSRWQS--RAF------PQLGGRPGPEGE----- 64
              |:||  ::  | |||       :| |: | ||  |:        ||     ||||:
Db        456 QDELQSKQSKGLEERYHRLSPPLPHEE-RAQSPPRSLATEEEPPQ-----GPEGQPEWKE 509

Qy         65 ---------GSLESQPPPLQTQA-CPESSCLREGEKGQNGDDSSAGGDFPPPAE----VE 110
                       ||  |    : ||  | ::| :  |:     ::   | :     |    |
Db        510 AEELGEDSAASLSLQLSLQREQAPSPPAACEKGKEQHSQAEELGPGQEEAEDPEEKVAVS 569

Qy        111 PTP-------EAELLAQPCHDSEASKLGAPAAGGEEEWGQQQRQLGKKK----------- 152
              |||          | :| |   |||: | |      :   | || | : |
Db        570 PTPPVSPEVRSTEPVAPPEQLSEAA-LKAMEEAVAQVLEQDQRHLLESKQEKMQQLREKL 628

Qy        153 -----------HRRRPSKKKRHWKPYYKLTWEEKKKFDEKQSLRASRIRAEMFAKGQPVA 201
                         |:::        :   |   ||: :  |::| | | :||:: :  |
Db        629 CQEEEEEILRLHQQKEQSLSSLRERLQKAIEEEEARMREEESQRLSWLRAQVQSSTQA-- 686

Qy        202 PYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSD--------DTSDDDFMEEGGEEDGGS- 252
                    | ||   : :  |   |   :|        :  :   :|:  ||   |
Db        687 ----------DEDQIRAEQEASLQKLREELESQQKAERASLEQKNRQMLEQLKEEIEASE 736

Qy        253 ----------------------DGMGGDGSEFLQRDFSETYERY-------HTE---SLQ 280
                                   :|   :    |::: |   ||         | |   |||
Db        737 KSEQAALNAAKEKALQQLREQLEGERKEAVATLEKEHSAELERLCSSLEAKHREVVSSLQ 796

Qy        281 NMSKQELIKEYLELEKCLSRMEDENNRLRLESKRLGGDDARVREL------------ELE 328
                ::   ||  :|:||| ::|   :|:  :|  : | :  :  |             |
Db        797 KKIQEAQQKEEAQLQKCLGQVE---HRVHQKSYHVAGYEHELSSLLREKRQEVEGEHERR 853

Qy        329 LDRLRAENLQLLT----ENELHRQQERAPL 354
              ||::: |: |::     : |   :::|| |
Db        854 LDKMKEEHQQVMAKAREQYEAEERKQRAEL 883


RESULT 12
Q95LS7
ID   Q95LS7      PRELIMINARY;      PRT;  560 AA.
AC   Q95LS7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical 63.1 kDa protein.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
RA   Terao K., Sugano S.;
RT   "Isolation of novel full-length cDNA clones from macaque testis cDNA
RT   libraries.";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB071115; BAB64509.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   560 AA;  63140 MW;  C3EF0D9DC4635E36 CRC64;

  Query Match             7.5%;  Score 142.5;  DB 6;  Length 560;
  Best Local Similarity   22.4%;  Pred. No. 0.023;
  Matches   88;  Conservative   55;  Mismatches  121;  Indels  129;  Gaps   19;

Qy         27 ELNPERPPGAEERVPEEDSRWQSRAFPQ----------LGGRPGPEGEGSLESQPPPLQ 75
              |  ||     ||   || |:  : | |             |  || ||    ::| |
Db         44 ESEPEEEEEEEEEEEEEASQGGTAADEQAKVPKELTAAEAAGEEGP-GEPGRPAKPQPE- 101
```

```
Qy          76 TQACPESSCLREGEKGQNGDDSSAGGDFP-------PPAEVEPTPEAELLAQPCHDSEA- 127
Db         102 ----PE-------------EPAEAGAEEPVQPKSGAGPEELDAEARAEELEQAAEGKEVR 144

Qy         128 -------SKLG------APAA-----GGEEEWGQQQRQLGKKKHRR----RPSKKKRHWK 165
Db         145 SQASLPLTRIGEEEAAAAPEAETERVEGEEEDEEETRRDGAESEGRAGEGRPAKSQEEGK 204

Qy         166 P------YYKLTWEEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPD 219
Db         205 PLGGRDEFEDLEWSE-----EVQKLQEQQLRSDLLDQYR--------SLLMERNRSQ--- 248

Qy         220 LKTGLY---------SKRAAAKSDDTSDDDFMEEGGEEDGG---SDGMGGDGSEFLQRDF 267
Db         249 -RYNLYLQHKIFEALRKKKGLEAAEVPDRGAQAEAPEKEQAYLRHLGM----LEDLKKQQ 303

Qy         268 SETYERYHTESLQNMSKQELIKEYLELEKCLSRMEDENNRLRLESKRL-----------G 316
Db         304 ADDLQWYH---------QELGQLKQQCQEKLSRVEKEWRRFQALKKQVVMQAMGSCRMRG 354

Qy         317 GDDARVREL----------ELELDRLRAENLQL 339
Db         355 GRQAALREVEQILALEDKKEKEMSAVRLENVQL 387
```

```
RESULT 13
Q9W475
ID   Q9W475      PRELIMINARY;      PRT;  1132 AA.
AC   Q9W475;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CG3108 protein.
GN   CG3108.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003435; AAF46083.2; -.
DR   HSSP; P00730; 2CTC.
DR   FlyBase; FBgn0029807; CG3108.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
SQ   SEQUENCE   1132 AA;  125335 MW;  09E481A1C6D7E526 CRC64;
```

```
  Query Match             7.4%;  Score 141.5;  DB 5;  Length 1132;
  Best Local Similarity   22.3%;  Pred. No. 0.062;
  Matches   80;  Conservative   57;  Mismatches  138;  Indels   83;  Gaps   17;

Qy           4 PFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPGPEG 63
Db         192 PLNDELPEDEESPAATESAV--EELEKESEAAMDDQVPEESEIQPEQVQP---GEYQSES 246

Qy          64 EG-------SLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAE 116
Db         247 DGEQAETKPEIEAQP---EVEAQPEAEAQPEAE--------------PQLEVEPQPEVE 288

Qy         117 LLAQPCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKK 176
Db         289 --SQPEVESQPEVEAQPEVEPQSEVESQ------------PEAESHSEPE---TQAEVE 330

Qy         177 KFDEKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDT 236
Db         331 AQPEVESLPEAESQPE--AESQP--------------EREPEVE----AEKISDNEVDT 369
```

```
Qy        237 SDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEK 296
              ::   ||   | | ||:       : : :|  ::  || |::  :|  : | :| |:
Db        370 TEASLMETLVE--GIEDGLTAAMDNLVPEELAEASDKQETE-LESEDQQSPVTEAIE-EQ 425

Qy        297 CLSRMEDENNRLRLESKRLGGDDARVRELELELDRLRAENLQ-LLTENELHRQQERAP 353
              :  :| |  :   | :::   |       | | |  :  | : :    | | : | ||
Db        426 AVPEIEQEKEK---EPEQITLAD------ETEQDSAQPSNEEPVEIAPEQHTEAEIAP 474


RESULT 14
O76153
ID   O76153      PRELIMINARY;      PRT;   443 AA.
AC   O76153;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Rsp60.
OS   Periplaneta americana (American cockroach).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Orthopteroidea; Dictyoptera; Blattaria; Blattoidea;
OC   Blattidae; Periplaneta.
OX   NCBI_TaxID=6978;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20193551; PubMed=10727895;
RA   Arai T., Kubo T., Natori S.;
RT   "Identification, characterization and cDNA cloning of two novel
RT   proteins secreted into the external space of the regenerating leg of
RT   Periplaneta americana.";
RL   Insect Biochem. Mol. Biol. 30:287-295(2000).
DR   EMBL; AB012307; BAA32795.1; -.
SQ   SEQUENCE   443 AA;  47987 MW;  91AFF646B161C99B CRC64;

  Query Match             7.4%;  Score 140.5;  DB 5;  Length 443;
  Best Local Similarity   20.4%;  Pred. No. 0.024;
  Matches   80;  Conservative   66;  Mismatches  140;  Indels  107;  Gaps   17;

Qy          4 PFLSEYQHQPQTSNCTGAAA--------VQEELNPERPPGAEERVPEEDSRWQSRAFPQL 55
              |  :|   :|  :  |  ||          ||: |   |  ||  | ||:
Db         47 PVAAEVPKEPVAAAATTVAADVPKEDEKPAEEVPPPAAPVVEEMKPAEDA---------- 96

Qy         56 GGRPGPEGEGSLESQPPPLQTQACPE---SSCLREGEKGQNGD-----DSSAGGDFPPPA 107
                 |  |   : : | :  :| ||        ::||  |:|      ::::  |   |
Db         97 ---PSPAPEAPAQPEVPAV-PEAVPEVAQDETNKKGEVNQDGSSDLKTEATSSQDVAKPV 152

Qy        108 E-----------------VEPTPEAELLAQPCHDSEASKLGAPAAGGEEEWGQQQRQLGK 150
              |                 |||  | ::      |  |      |    |||   :     |
Db        153 EEKVALNAPKVGEEPAEKVEPAEEKKVEEVKVQDQAAPTTAEPPKAEEEEKPAKDE---K 209

Qy        151 KKHRRRP-SKKKRHWKPYYKLTWEEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNTTQFL 209
              :    | |::||   |      |||||  ||::     :||
Db        210 VEEAAAPVSRRKRETAP----KEEEKKKPTEKKA------KAE---------------- 242

Qy        210 MDDHDQEEPDLKTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSE 269
              |  ||||     :  :   : :   :| ::: :::             :::|  |
Db        243 -DVKVQEEPKQDVKVQEEEVKVQDEPKQEDVQVQDDAKKE----------EVKVEQDAKE 291

Qy        270 TYERYHTESLQNMSKQELIKEYLELEKCLSRMEDENNRLRLESKRLGGD---DARVRELE 326
              :   |  |:   |:: :: :       :::||   ::||  ::  |   : :| | :
Db        292 EEVKVQEEVKQDAKAQDVKQDDV-------KVQDE---VKLEEVKVESDAKEEVKVEEAK 341

Qy        327 LELDRLRAENLQLLTENELHRQ---QERAPLSK 356
              :| |    |:::: ::|: :      |  |:||
Db        342 VEQDAKVQEDVKV--QDEVKQDAAASEVNPVSK 372


RESULT 15
Q8TXA4
ID   Q8TXA4      PRELIMINARY;      PRT;   609 AA.
AC   Q8TXA4;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Uncharacterized protein.
GN   MK0771.
OS   Methanopyrus kandleri.
OC   Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC   Methanopyrus.
OX   NCBI_TaxID=2320;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AV19 / DSM 6324 / JCM 9639;
RX   MEDLINE=21927647; PubMed=11930014;
RA   Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,
RA   Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA   Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA   Malykh A.G., Koonin E.V., Kozyavkin S.A.;
RT   "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT   and monophyly of archaeal methanogens.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
DR   EMBL; AE010369; AAM01985.1; -.
DR   HSSP; P04268; 1IC2.
DR   InterPro; IPR002017; Spectrin.
KW   Complete proteome.
SQ   SEQUENCE   609 AA;  69552 MW;  AB10C9780DC5AD78 CRC64;

  Query Match             7.3%;  Score 140;  DB 17;  Length 609;
  Best Local Similarity   22.0%;  Pred. No. 0.039;
  Matches   80;  Conservative   66;  Mismatches  147;  Indels   70;  Gaps   17;

Qy         31 ERPPGAEER--------VPEEDSRWQSRAFPQLGGRPGPEGEGS-----LESQPPPLQTQ 77
              :|  |  ||         |||          |:  |||| | :|:     |: :||
Db          5 KRKGGKRERPEIDLKLKPPEEGEGAPKLKLPKPEGRPG-ESKGAEAVPKLKLKPP----- 58

Qy         78 ACPESSCLREGEKGQNGDDSSAGGDFPPPAEVE-PTPEAELLAQPCHDSEASKLGAPAAG 136
                     :  ||  :|::      :  ||||:: | |  |  |    ::|  :| |
Db         59 --------KPSEKPPSGEEEEKKKEERPPAEIKPPKPPEERTAPSNLEAELERLKAENKK 110

Qy        137 GEE---EW--------GQQQRQLGK-KKHRRRPSKKKRHWKPYYKLTWEEKKKFDE---- 180
              |   ||         |:: |   : |: :    |:::    | |:: : |:| ::
Db        111 LREELDEWRNKAKSAMGERDRLRSEIKRLKEELEKQEKELDKYIKISKQLKEKLEKAKRE 170

Qy        181 ----KQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDT 236
                  |:     | | |  |        ||  :  ::|   :    |   |  |:   | ::
Db        171 SEELKEKAEEYRERYEKIA-----GKYNELKSKLEDLSDQNRRLAENL--KKLKEKYNEI 223

Qy        237 SDD-DFMEEGGEEDGGSDGMGGDGSEFLQRDFSET-YER----YHTESLQNMS---KQEL 287
              :: | ::|  :| |        |     ||     |  ||        |:|:|  :   ::::
Db        224 KEERDRLKEETKEVGKLK----DQLAKLQSKLKEVKSERDDLANEVEALRNENEKLRKKI 279

Qy        288 IKEYLELEKCLSRMEDENNRLRLESKRLGGDDARVRELELELDRLRAENLQLLTENELHR 347
              |   ||     :::|   :|    : :|     ::  : |: :||    :|  ::|: |
Db        280 DKLKSELSNLQKKLKDREKKLEKARQHIGKLREEIKRRDEEIRKLRKAQSKL--KDEIKR 337

Qy        348 QQE 350
              :|
Db        338 YEE 340


Search completed: February  5, 2004, 13:31:02
Job time : 112 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 5, 2004, 13:38:50 ; Search time 2658 Seconds
(without alignments)
3282.659 Million cell updates/sec

Title: US-09-972-758A-2
Perfect score: 1910
Sequence: 1 MAEPFLSEYQHQPQTSNCTG.....LTENELHRQQERAPLSKFGD 359

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09972758/runat_05022004_095003_1203/app_query.fasta_1.519
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09972758_@CGN_1_1_2810_@runat_05022004_095003_1203 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1414.5	74.1	894	14	CD518239	CD518239 AGENCOURT
	2	1350	70.7	1201	9	AL552998	AL552998 AL552998
	3	1327.5	69.5	903	13	BQ221641	BQ221641 AGENCOURT
c	4	1307	68.4	750	13	BX112898	BX112898 BX112898
	5	1266.5	66.3	771	12	BI871190	BI871190 603395081
	6	1231	64.5	982	10	BF984049	BF984049 602306989
	7	1163.5	60.9	797	12	BG819109	BG819109 602781202
	8	1142.5	59.8	852	13	BU170664	BU170664 AGENCOURT
	9	1137.5	59.6	830	13	BU526446	BU526446 AGENCOURT
	10	1129	59.1	750	9	AV726891	AV726891 AV726891
	11	1120	58.6	852	12	BI091005	BI091005 602853667
c	12	1120	58.6	1078	9	AL578242	AL578242 AL578242
	13	1111.5	58.2	871	10	BG291150	BG291150 602387182
	14	1074	56.2	1294	12	BM474968	BM474968 AGENCOURT
	15	1059	55.4	683	14	CA430432	CA430432 UI-H-DF0-
	16	1056.5	55.3	1080	12	BM474949	BM474949 AGENCOURT
	17	1042	54.6	949	10	BG165450	BG165450 602345830
c	18	996	52.1	1201	9	AL576112	AL576112 AL576112
	19	963	50.4	557	12	BM129736	BM129736 if22e01.y
	20	953	49.9	762	10	BF167480	BF167480 601775078
	21	946.5	49.6	745	10	BE547281	BE547281 601073706
	22	933	48.8	907	10	BF181535	BF181535 601808833
	23	932.5	48.8	666	12	BG864490	BG864490 602798512
	24	909	47.6	810	10	BF607249	BF607249 MY1_00019
	25	905	47.4	999	12	BI411874	BI411874 602966526
	26	900	47.1	933	10	BF674587	BF674587 602137805
	27	893	46.8	535	14	CA865829	CA865829 ir42c06.y
	28	879	46.0	563	9	AW665096	AW665096 hi99d03.x
	29	879	46.0	843	10	BF026064	BF026064 601670105
	30	868.5	45.5	807	13	BU461116	BU461116 603368286
	31	860	45.0	638	9	AA546538	AA546538 vk65f09.s
	32	858	44.9	591	13	BU471258	BU471258 603762160
	33	857.5	44.9	765	13	BU489324	BU489324 604129287
	34	857	44.9	727	13	BU400268	BU400268 604138634
	35	852.5	44.6	920	13	BU249231	BU249231 603779322
	36	851	44.6	547	9	AI796944	AI796944 we25b03.x
	37	850	44.5	1072	9	AL555845	AL555845 AL555845
	38	841	44.0	482	9	AI948609	AI948609 wq07g11.x
	39	841	44.0	566	10	AW962738	AW962738 EST374811
	40	829.5	43.4	871	13	BU314320	BU314320 603543563
	41	826.5	43.3	656	13	BU287394	BU287394 604163285
	42	826.5	43.3	911	13	BU288715	BU288715 603606177
	43	822	43.0	462	9	AV728223	AV728223 AV728223
	44	814.5	42.6	741	13	BU374772	BU374772 603590650
	45	812	42.5	907	12	BI412550	BI412550 602990856

ALIGNMENTS

RESULT 1
CD518239
LOCUS CD518239 894 bp mRNA linear EST 06-JUN-2003
DEFINITION AGENCOURT_14375843 NIH_MGC_181 Homo sapiens cDNA clone IMAGE:30407314 5', mRNA sequence.
ACCESSION CD518239
VERSION CD518239.1 GI:31449957
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 894)

```
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM488  row: h  column: 11
            High quality sequence stop: 602.
FEATURES              Location/Qualifiers
     source           1. .894
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:30407314"
                      /tissue_type="White Matter"
                      /dev_stage="Unknown"
                      /lab_host="DH10B-Ton A ( T1 and T5 phage resistances)"
                      /clone_lib="NIH_MGC_181"
                      /note="Vector: pCMV-SPORT6.1; Site_1: NotI; Site_2: EcoRV
                      (destroyed); Library is oligo-dT primed and directionally
                      cloned (EcoRV site is destroyed upon cloning). Average
                      insert size 1.42 kb. Library was constructed by
                      (Invitrogen). Note: this is a NIH_MGC Library."
BASE COUNT      218 a    244 c    314 g    114 t      4 others
ORIGIN

Alignment Scores:
Pred. No.:                2.31e-74        Length:          894
Score:                    1414.50         Matches:         273
Percent Similarity:       95.49%          Conservative:    2
Best Local Similarity:    94.79%          Mismatches:      9
Query Match:              74.06%          Indels:          4
DB:                       14              Gaps:            1

US-09-972-758A-2 (1-359) x CD518239 (1-894)

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 60

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaProProGlnLeuGlyGlyArgProGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 120

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 180

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 240

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 300

Qy        121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 360

Qy        141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 420

Qy        161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
              ||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||
Db        421 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGNGAAGAGAAGAAAAAGTTCGACGAG 480

Qy        181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 540

Qy        201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 600

Qy        221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ANAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 660

Qy        241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260
              ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        661 TTCATGGAAGAAGGNGGTGAGGGAGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 720

Qy        261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 780

Qy        281 AsnMetSerLysGlnGluLeuIle-LysGluTyrLeuGluLeu--GluLysCys---Leu 298
              ||||||   :::|||||||||||| :::|||||||||||     |||||||||  . |||
Db        781 AACATGAGGCAGCAGGAGCTCATTCAGGGAGTACCTGGAAACTGGGAGAAGTGCCCTCTT 840

Qy        299 SerArgMetGluAspGlu 304
                    |||||||||
Db        841 CGCGCATGGGAGGACGAA 858

RESULT 2
AL552998
LOCUS       AL552998               1201 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL552998 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI072YL06 5-PRIME, mRNA sequence.
ACCESSION   AL552998
VERSION     AL552998.2  GI:31274812
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12892417.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 8422.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI072DF03QP1&cluster=8422.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI072DF03QP1.
FEATURES              Location/Qualifiers
     source           1. .1201
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="CS0DI072YL06"
                      /tissue_type="PLACENTA COT 25-NORMALIZED"
                      /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                      /note="1st strand cDNA was primed with a NotI-oligo(dT)
                      primer. Five prime end enriched, double-strand cDNA was
                      digested with Not I and cloned into the Not I and EcoR V
                      sites of the pCMVSPORT 6 vector. Library was normalized."
```

```
BASE COUNT      295 a    280 c    360 g    200 t     66 others
ORIGIN

Alignment Scores:
Pred. No.:              1.88e-70        Length:         1201
Score:                  1350.00         Matches:        262
Percent Similarity:     95.99%          Conservative:   1
Best Local Similarity:  95.62%          Mismatches:     10
Query Match:            70.68%          Indels:         4
DB:                     9               Gaps:           1

US-09-972-758A-2 (1-359) x AL552998 (1-1201)

Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        202 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGK 261

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        262 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 321

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        322 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 381

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        382 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 441

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        442 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 501

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        502 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 561

Qy        121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db        562 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGG-GGCGAAGAGGAG 620

Qy        141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 680

Qy        161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 740

Qy        181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        741 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 800

Qy        201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        801 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 860

Qy        221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 920

Qy        241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260
              |||||||||   ||||||||||||||||||||    ||||||||||||||||||||:::
Db        921 TTCATGGAARAAGGGGGTGAGGAGGATGGGGGMRC-GATGGGATGGGAGGGGACGGMGCS 979

Qy        261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyr 274
                 |||                 ||||||   |||||||||
Db        980 AGTTTY---TSMGCGGGACTTYYCGAGACGTA-GAGCGGTAC 1017

RESULT 3
BQ221641
LOCUS       BQ221641                 903 bp    mRNA    linear   EST 02-MAY-2002
DEFINITION  AGENCOURT_7552882 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6058110
            5', mRNA sequence.
ACCESSION   BQ221641
VERSION     BQ221641.1  GI:20403041
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 903)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP/Gazdar
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13322  row: k  column: 07
            High quality sequence stop: 644.
FEATURES             Location/Qualifiers
     source          1. .903
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6058110"
                     /tissue_type="large cell carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_68"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.8 kb.   Library constructed by Life
                     Technologies. "
BASE COUNT      217 a    259 c    309 g    115 t      3 others
ORIGIN

Alignment Scores:
Pred. No.:              3.15e-69        Length:         903
Score:                  1327.50         Matches:        251
Percent Similarity:     97.68%          Conservative:   2
Best Local Similarity:  96.91%          Mismatches:     4
Query Match:            69.50%          Indels:         2
DB:                     13              Gaps:           1

US-09-972-758A-2 (1-359) x BQ221641 (1-903)

Qy         51 AlaProProGlnLeuGlyGlyArgProGlyProGluGlyGluGlySerLeuGluSerGln 70
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCGTTCCCCCAGTTGGGTGGCCGTCCGGGGCCGGAGGGGGAAGGGAGCCTGGAATCCCAA 60

Qy         71 ProProProLeuGlnThrGlnAlaCysProGluSerSerCysLeuArgGluGlyGluLys 90
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCACCTCCCTTGCAGACCCAGGCCTGTCCAGAATCTAGCTGCCTGAGAGAGGGCGAGAAG 120

Qy         91 GlyGlnAsnGlyAspAspSerSerAlaGlyGlyAspPheProProProAlaGluValGlu 110
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCCAGAATGGGGACGACTCGTCCGCTGGCGGCGACTTCCCGCCGCCGGCAGAAGTGGAA 180

Qy        111 ProThrProGluAlaGluLeuLeuAlaGlnProCysHisAspSerGluAlaSerLysLeu 130
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCGACGCCCGAGGCCGAGCTGCTCGCCCAGCCTTGTCATGACTCCGAGGCCAGTAAGTTG 240

Qy        131 GlyAlaProAlaAlaGlyGlyGluGluGluTrpGlyGlnGlnGlnArgGlnLeuGlyLys 150
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGGGCTCCTGCCGCAGGGGGCGAAGAGGAGTGGGGACAGCAGCAGAGACAGCTGGGGAAG 300

Qy        151 LysLysHisArgArgArgProSerLysLysLysArgHisTrpLysProTyrTyrLysLeu 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAAAAACATAGGAGACGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTACTACAAGCTG 360
```

```
Qy        171 ThrTrpGluGluLysLysLysPheAspGluLysGlnSerLeuArgAlaSerArgIleArg 190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACCTGGGAAGAGAAGAAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGA 420

Qy        191 AlaGluMetPheAlaLysGlyGlnProValAlaProTyrAsnThrThrGlnPheLeuMet 210
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GCCGAGATGTTCGCCAAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTCATG 480

Qy        211 AspAspHisAspGlnGluGluProAspLeuLysThrGlyLeuTyrSerLysArgAlaAla 230
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GATGATCACGACCAGGAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCC 540

Qy        231 AlaLysSerAspAspThrSerAspAspAspPheMetGluGluGlyGlyGluGluAspGly 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCAAATCCGACGACACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGG 600

Qy        251 GlySerAspGlyMetGlyGlyAspGlySerGluPheLeuGlnArgAspPheSerGluThr 270
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCAGCGATGGGATGGGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACG 660

Qy        271 TyrGluArgTyrHisThr-GluSerLeuGlnAsnMetSerLysGlnGluLeuIleLysGl 290
              ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        661 TACGAGCGGTACCACACGGNAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAAGA 720

Qy        290 uTyrLeuGluLeuGluLysCys---LeuSerArgMetGluAspGluAsnAsnArg 307
              |||||||  ||||||||||||  |||:::      :::|||||||||||||||
Db        721 GTACCTGGGACTGGAGAAGTGCCCTCTCGCCGCATGGAAGGACGAGAACAACCGG 775

RESULT 4
BX112898/c
LOCUS       BX112898                750 bp    mRNA    linear   EST 07-FEB-2003
DEFINITION  BX112898 Soares melanocyte 2NbHM Homo sapiens cDNA clone
            IMAGp998A05595 ; IMAGE:270604, mRNA sequence.
ACCESSION   BX112898
VERSION     BX112898.1  GI:27837920
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 750)
  AUTHORS   Ebert,L., Heil,O., Hennig,S., Neubert,P., Partsch,E., Peters,M.,
            Radelof,U., Schneider,D. and Korn,B.
  TITLE     Human UnigeneSet - RZPD3
  JOURNAL   Unpublished
COMMENT     Contact: Ina Rolfs
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
            RZPD; IMAGp998A05595.
            RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
            Human UnigeneSet - RZPD3 (RZPDLIB No.972)
            http://www.rzpd.de/CloneCards/cgi-
            bin/showLib.pl.cgi/response?libNo=972 Contact: Ina Rolfs
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Heubnerweg 6, D-14059 Berlin, Germany
            Tel: +49 30 32639 101
            Fax: +49 30 32639 111
            www.rzpd.de
            This clone is available royalty-free from RZPD;
            contact RZPD (clone@rzpd.de) for further information. Seq primer:
            M13r, Primer sequence: TTTCACACAGGAAACAGCTATGAC.
FEATURES              Location/Qualifiers
     source           1. .750
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGp998A05595 ; IMAGE:270604"
                      /sex="Male"
                      /tissue_type="melanocyte"
                      /lab_host="DH10B (ampicillin resistant)"
                      /clone_lib="Soares melanocyte 2NbHM"
                      /note="Vector: pT7T3D (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCAGTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library constructed by Bento Soares and
                      M.Fatima Bonaldo. RNA from normal foreskin melanocytes
                      (FS374) was kindly provided by Dr. Anthony P. Albino."
BASE COUNT       92 a    265 c    200 g    193 t
ORIGIN

Alignment Scores:
Pred. No.:              4.37e-68        Length:         750
Score:                  1307.00         Matches:        249
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.60%          Mismatches:     0
Query Match:            68.43%          Indels:         0
DB:                     13              Gaps:           0

US-09-972-758A-2 (1-359) x BX112898 (1-750)

Qy        107 AlaGluValGluProThrProGluAlaGluLeuLeuAlaGlnProCysHisAspSerGlu 126
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        750 GCAGAAGTGAAGCCGACGCCCGAGGCCGAGCTGCTCGCCCAGCCTTGTCATGACTCCGAG 691

Qy        127 AlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGluTrpGlyGlnGlnGlnArg 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        690 GCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAGTGGGGACAGCAGCAGAGA 631

Qy        147 GlnLeuGlyLysLysLysHisArgArgArgProSerLysLysLysArgHisTrpLysPro 166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        630 CAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAGAAGCGGCATTGGAAACCG 571

Qy        167 TyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGluLysGlnSerLeuArgAla 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        570 TACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAGAAACAGAGCCTTCGAGCT 511

Qy        187 SerArgIleArgAlaGluMetPheAlaLysGlyGlnProValAlaProTyrAsnThrThr 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        510 TCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTCGCGCCCTATAACACCACG 451

Qy        207 GlnPheLeuMetAspAspHisAspGlnGluGluProAspLeuLysThrGlyLeuTyrSer 226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        450 CAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGACCTCAAAACCGGCCTGTACTCC 391

Qy        227 LysArgAlaAlaAlaLysSerAspAspThrSerAspAspAspPheMetGluGluGlyGly 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        390 AAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGACTTCATGGAAGAAGGGGGT 331

Qy        247 GluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySerGluPheLeuGlnArgAsp 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        330 GAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGCGAGTTTCTGCAGCGGGAC 271

Qy        267 PheSerGluThrTyrGluArgTyrHisThrGluSerLeuGlnAsnMetSerLysGlnGlu 286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        270 TTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAGAACATGAGCAAGCAGGAG 211

Qy        287 LeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArgMetGluAspGluAsnAsn 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        210 CTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGCATGGAGGACGAGAACAAC 151

Qy        307 ArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAlaArgValArgGluLeuGlu 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        150 CGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCGCGTGTGCGGGAGCTGGAG 91

Qy        327 LeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnGluLeuHis 346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         90 CTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTGACCGAGAACGAACTGCAC 31

Qy        347 ArgGlnGlnGluArgAlaProLeuSerLys 356
              |||||||||||||||||||||||||||||
Db         30 CGGCAGCAGGAGCGAGCGCCGCTTTCCAAG 1
```

```
RESULT 5
BI871190
LOCUS       BI871190                771 bp    mRNA    linear   EST 11-OCT-2001
DEFINITION  603395081F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:5404724 5',
            mRNA sequence.
ACCESSION   BI871190
VERSION     BI871190.1  GI:16044865
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 771)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12032  row: b  column: 21
            High quality sequence stop: 733.
FEATURES             Location/Qualifiers
     source          1. .771
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5404724"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_90"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.7 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      187 a    213 c    273 g     98 t
ORIGIN

Alignment Scores:
Pred. No.:              1.09e-65      Length:         771
Score:                  1266.50       Matches:        248
Percent Similarity:     96.51%        Conservative:   1
Best Local Similarity:  96.12%        Mismatches:     5
Query Match:            66.31%        Indels:         5
DB:                     12            Gaps:           1

US-09-972-758A-2 (1-359) x BI871190 (1-771)

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 62

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 122

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 182

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 242

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 302

Qy        121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 362

Qy        141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 422

Qy        161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 482

Qy        181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 542

Qy        201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 602

Qy        221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240
              |||||||||||||||||||||||||||   |||||||||||||||||||||||||||||||
Db        603 AAAACCGGCCTGTACTCCAAGCGGGCCGC-GCCAAATCCGACGACACCAGCGATGACGAC 661

Qy        241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGly---GlyAspGly 259
              |||||||||||||||||||||||||||||||||||||||||||   |||   |||
Db        662 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGAATGGGAAGGGGACGGCC 721

Qy        260 SerGlu-PheLeuGlnArgAsp-PheSerGlu-ThrTyrGluArgTyr 274
              |||||| |||   ||||||::: ||||||||| ||||||||||||||
Db        722 AGCGAGGTTCTGCCAGCGGGAACTTCTCGGAGAACGTACGAGCGGTAC 769

RESULT 6
BF984049
LOCUS       BF984049                982 bp    mRNA    linear   EST 23-JAN-2001
DEFINITION  602306989F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4398238 5',
            mRNA sequence.
ACCESSION   BF984049
VERSION     BF984049.1  GI:12386861
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 982)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10100  row: a  column: 23
            High quality sequence stop: 691.
FEATURES             Location/Qualifiers
     source          1. .982
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4398238"
                     /tissue_type="duodenal adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_88"
                     /note="Organ: small intestine; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
                     oligo-dT primed. Average insert size 1.767 kb. Library
                     enriched for full-length clones and constructed by Life
```

```
                    Technologies. Note: this is a NIH_MGC Library."
BASE COUNT      261 a    235 c    320 g    166 t
ORIGIN

Alignment Scores:
Pred. No.:              1.66e-63        Length:         982
Score:                  1231.00         Matches:        238
Percent Similarity:     99.17%          Conservative:   0
Best Local Similarity:  99.17%          Mismatches:     2
Query Match:            64.45%          Indels:         1
DB:                     10              Gaps:           0

US-09-972-758A-2 (1-359) x BF984049 (1-982)

Qy        120 GlnProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGlu 139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          8 CAGCCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAG 67

Qy        140 GluTrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLys 159
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         68 GAGTGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAG 127

Qy        160 LysLysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAsp 179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        128 AAGAAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGAC 187

Qy        180 GluLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnPro 199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        188 GAGAAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCG 247

Qy        200 ValAlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAsp 219
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        248 GTCGCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGAT 307

Qy        220 LeuLysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAsp 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        308 CTCAAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGAC 367

Qy        240 AspPheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGly 259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        368 GACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGC 427

Qy        260 SerGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeu 279
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        428 AGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTG 487

Qy        280 GlnAsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSer 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        488 CAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCG 547

Qy        300 ArgMetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAsp 319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        548 CGCATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGAC 607

Qy        320 AlaArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeu 339
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        608 GCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTG 667

Qy        340 LeuThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359
              ||||||||||||||||||||||||||||||||||||   |||   ||||||||||||
Db        668 CTGACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCG-CTTCCCAAGTTTGGAGAC 726

RESULT 7
BG819109
LOCUS       BG819109                797 bp    mRNA    linear   EST 22-MAY-2001
DEFINITION  602781202F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4932025
            5', mRNA sequence.
ACCESSION   BG819109
VERSION     BG819109.1  GI:14166696
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 797)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10857  row: c  column: 02
            High quality sequence stop: 797.
FEATURES             Location/Qualifiers
     source          1. .797
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4932025"
                     /tissue_type="anaplastic oligodendroglioma with 1p/19q
                     loss"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Brn67"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2.3 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      207 a    223 c    248 g    119 t
ORIGIN

Alignment Scores:
Pred. No.:              1.34e-59        Length:         797
Score:                  1163.50         Matches:        232
Percent Similarity:     95.12%          Conservative:   2
Best Local Similarity:  94.31%          Mismatches:     8
Query Match:            60.92%          Indels:         6
DB:                     12              Gaps:           1

US-09-972-758A-2 (1-359) x BG819109 (1-797)

Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         69 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 128

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        129 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 188

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        189 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 248

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        249 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 308

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        309 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 368

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        369 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 428

Qy        121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        429 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 488

Qy        141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        489 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 548

Qy        161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLys--LysPheAspG 180
              ||||||  ||||||||||||||||||||||||||||||||||:::  ||||||||||
Db        549 AAGCGGCA-TGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGCAACAAAGTTCGACG 607

Qy        180 luLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProV 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        608 AGAAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGG 667

Qy        200 alAlaProTyrAsnThrThrGlnPheLeuMet-AspAspHisAspGlnGluGluProAsp 219
              |||||   |||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        668 TCGCGCGCTATAACACCACGCAGTTCCTCATGGGATGATCACGACCAGGAGGAGCCGGAT 727

Qy        220 LeuLysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAsp 239
              ||||||||||||   ||||||||||||   ||||||  :::           ||||||
Db        728 CTCAAAACCGGCTG-TACTCCAAGCGGGGCGCGGCA---AATCCGAGACAACAGGATGAC 783

Qy        240 AspPheMetGlu 243
              ||||||||||||
Db        784 GACTTCATGGAA 795

RESULT 8
BU170664
LOCUS       BU170664                852 bp    mRNA    linear   EST 04-SEP-2002
DEFINITION  AGENCOURT_7753860 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6094335
            5', mRNA sequence.
ACCESSION   BU170664
VERSION     BU170664.1  GI:22684648
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 852)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13364  row: p  column: 16
            High quality sequence stop: 624.
FEATURES             Location/Qualifiers
     source          1. .852
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6094335"
                     /tissue_type="epithelioid carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_70"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.1 kb.   Library constructed by Life
                     Technologies."
BASE COUNT      202 a    254 c    265 g    131 t
ORIGIN

Alignment Scores:
Pred. No.:              2.45e-58       Length:         852
Score:                  1142.50        Matches:        212
Percent Similarity:     97.27%         Conservative:   2
Best Local Similarity:  96.36%         Mismatches:     5
Query Match:            59.82%         Indels:         1
DB:                     13             Gaps:           1

US-09-972-758A-2 (1-359) x BU170664 (1-852)

Qy         18 CysThrGlyAlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGlu 37
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 TGTACAGGTGCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAG 62

Qy         38 GluArgValProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGly 57
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||||
Db         63 GAGCGGGTGCCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCACTTGGGTGGC 122

Qy         58 ArgProGlyProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGln 77
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 CGTCCGGGGCCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAG 182

Qy         78 AlaCysProGluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSer 97
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 GCCTGTCCAGAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCG 242

Qy         98 SerAlaGlyGlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeu 117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 TCCGCTGGCGGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTG 302

Qy        118 LeuAlaGlnProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGly 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 CTCGCCCAGCCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGC 362

Qy        138 GluGluGluTrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgPro 157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 GAAGAGGAGTGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCG 422

Qy        158 SerLysLysLysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLys 177
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 TCCAAGAAGAAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAG 482

Qy        178 PheAspGluLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGly 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 TTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGC 542

Qy        198 GlnProValAlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGlu 217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 CAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAG 602

Qy        218 ProAspLeuLysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSer---AspAspThr 236
              ||||||||||||||||||||||||||:::|||           :::   |||||||||
Db        603 CCGGATCTCAAAACCGGCCTGTACTCCCAGCGGGGCCGCCGCCCAAATCCCGACGACACC 662

RESULT 9
BU526446
LOCUS       BU526446                830 bp    mRNA    linear   EST 13-SEP-2002
DEFINITION  AGENCOURT_10096655 NIH_MGC_144 Mus musculus cDNA clone
            IMAGE:6535344 5', mRNA sequence.
ACCESSION   BU526446
VERSION     BU526446.1  GI:22836887
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 830)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein
             cDNA Library Preparation: Michael Brownstein Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2694  row: o  column: 24
            High quality sequence stop: 559.
```

```
FEATURES             Location/Qualifiers
     source          1. .830
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6535344"
                     /lab_host="DH10B (T1-phage-resistant)"
                     /clone_lib="NIH_MGC_144"
                     /note="Organ: Brain; Vector: pDNR-LIB; Site_1: SfiI
                     (ggccattatggcc); Site_2: SfiI (ggccgcctcggcc); cDNA made
                     by oligo-dT priming and directionally cloned. 5' and 3'
                     adaptors were used in cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Full-length
                     enriched library was constructed using the Clontech
                     Creator SMART kit and size-selected to contain the 0.2-0.5
                     kb size fraction (other fractions present in NIH_MGC_143).
                     Library created in the laboratory of M. Brownstein (NIMH,
                     NIH). Note: this is a NIH_MGC Library."
BASE COUNT      212 a    230 c    279 g    108 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:                4.72e-58       Length:          830
Score:                    1137.50        Matches:         229
Percent Similarity:       87.23%         Conservative:    10
Best Local Similarity:    83.58%         Mismatches:      30
Query Match:              59.55%         Indels:          5
DB:                       13             Gaps:            2

US-09-972-758A-2 (1-359) x BU526446 (1-830)

Qy          83 SerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGlyGlyAsp 102
               |||   |||   :::|||||||||||||||||||:::|||   |||   ||||||
Db           4 AGCTCCCTGGAGAAGGGCGAGAAGGGCCAGAATGGGGAGGACTTATCCACTGGCGGT--- 60

Qy         103 ProProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGlnProCys 122
                   |||   ||||||   ||||||      |||:::     |||   ||||||
Db          61 GCCTCCCCGTCGGCGGAGGGAGAGCCGATGTCAGAGTCC------CTCGTGCAGCCAGGT 114

Qy         123 HisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGluTrpGly 142
               ||||||||||||||:::|||      ||||||||||||||||||||||   ||||||
Db         115 CATGACTCGGAGGCCACCAAGCAGGAGGCTCCTGCCGCCGGAGGCGAGGAGCCATGGGGA 174

Qy         143 GlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLysLysArg 162
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         175 CAGCAACAGAGACAGCTGGGCAAGAAAAAACATCGGAGACGCCCCTCAAAGAAGAAGCGG 234

Qy         163 HisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGluLysGln 182
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         235 CATTGGAAGCCCTACTACAAGCTGACTTGGGAGGAGAAGAAAAAGTTCGACGAGAAGCAG 294

Qy         183 SerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProValAlaPro 202
               ||||||||||||||||:::||||||||||||||||||||||||||||||||||||||||
Db         295 AGCCTGCGAGCTTCGCGGGTTCGAGCCGAGATGTTCGCCAAGGGCCAGCCAGTTGCGCCC 354

Qy         203 TyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeuLysThr 222
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         355 TATAACACCACGCAGTTCCTCATGGATGACCACGATCAGGAGGAGCCTGATCTCAAAACC 414

Qy         223 GlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAspPheMet 242
               |||||||||   ||||||||||||||||||||||||||||||:::|||||:::
Db         415 GGCCTTTACCCCAAGCGGGCAGCCGCCAAATCCGACGACACCAGCGATGAGGATTTTGTG 474

Qy         243 GluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySerGluPhe 262
               ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         475 GAAGAAGCTGGTGAGGAAGACGGAGGCAGCGATGGCATGGGAGGGGACGGCAGCGAGTTT 534

Qy         263 LeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGlnAsnMet 282
               ||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db         535 CTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACGCCGAGAGCCTGCAGAACATG 594

Qy         283 SerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArgMetGlu 302
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         595 AGCAAGCAGGAGCTCATCAAAGAGTACCTGGAGCTGGAGAAGTGCCTCTCCCGCCAGGAG 654

Qy         303 AspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAlaArgVal 322
               ||||||||||||||||||   |||   ||||||||||||||||||      ||||||
Db         655 GACGAAAATAACCGGCTGCCGCTGGGAAGCAAGCGGCTGGGAGGCGTCCACCCGCGAGTG 714

Qy         323 ArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeuThrGlu 342
               |||||||||   |||:::||||||||||||||||   |||   |||   ||||||||||||
Db         715 CGGGAGCTCNAGCTAAAGCTGGACCGGCTGCGCCCTGAGAAACTCCCACTGCTGACCGAG 774

Qy         343 -AsnGluLeu-HisArgGlnGlnGluArgAlaProLeu 354
                |||:::|||    ||||||||||||||:::   |||
Db         775 AAACAAACTGCCCCCGGCAGCAGGAACGGAGCGCCCTC 812

RESULT 10
AV726891
LOCUS       AV726891                750 bp    mRNA    linear   EST 17-OCT-2000
DEFINITION  AV726891 HTC Homo sapiens cDNA clone HTCAQE07 5', mRNA sequence.
ACCESSION   AV726891
VERSION     AV726891.1  GI:10836312
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 750)
  AUTHORS   Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X.,
            Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu
            ,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R.,
            Chen,J., Chen,Z. and Han,Z.
  TITLE     Homo sapiens cDNA HTC clones
  JOURNAL   Unpublished
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .750
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HTCAQE07"
                     /tissue_type="Hypothalamus"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /clone_lib="HTC"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      184 a    208 c    251 g     96 t     11 others
ORIGIN

Alignment Scores:
Pred. No.:                1.38e-57       Length:          750
Score:                    1129.00        Matches:         219
Percent Similarity:       85.94%         Conservative:    1
Best Local Similarity:    85.55%         Mismatches:      18
Query Match:              59.11%         Indels:          18
DB:                       9              Gaps:            1

US-09-972-758A-2 (1-359) x AV726891 (1-750)

Qy         101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           2 GGCGACTTCCCGCCGCCGGCAGAAGTGGAGCCGACGCCCGAGGCCGAGCTGCTCGCCCAG 61

Qy         121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          62 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 121
```

```
Qy        141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 181

Qy        161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
              ||||||||||||||||||||||||||||||||||||
Db        182 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAG-------------- 226

Qy        181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
                                                  ||||||||||||||||||
Db        227 ------------------------------------ATCGCCAAGGGCCAGCCGGTC 247

Qy        201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        248 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGACCTC 307

Qy        221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240
                  ||||||||||||||||||||||    ||||||||||||||||||||||||||||||||
Db        308 ANAACCGGCCTGTACTCCAAGCGGGCCGNCGCCAAATCCGACGACACCAGCGATGACGAC 367

Qy        241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        368 TTCATGGAAGAAGGGGGTGAGGAGGATGGNGGCAGCGATGGGATGGGAGGGGACGGCAGC 427

Qy        261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        428 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 487

Qy        281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArg 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        488 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTNTCGCGC 547

Qy        301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320
              ||||||||||||||||||||||    ||||||||||||||||||||||||||||||||||||
Db        548 ATGGAGGACGAGAACAACCGGCTGCNGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 607

Qy        321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340
                  ||||||||||||||    |||||||||   |||          :::       ||||||
Db        608 CCTGTGCGGGAGCTTGAGCCTGAGCTTGACCCGCTTGNNCGCCGAGACCCTTCACTGCTG 667

Qy        341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLys 356
              |||   ||||||         |||   ||||||||||||||    |||
Db        668 ACCGANAACGAACCTGACCCGCAGCANGAGCGAGCGCCGCTTCCAAAG 715

RESULT 11
BI091005
LOCUS       BI091005                852 bp    mRNA    linear   EST 20-JUN-2001
DEFINITION  602853667F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4995065 5',
            mRNA sequence.
ACCESSION   BI091005
VERSION     BI091005.1  GI:14509335
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 852)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Incyte Genomics, Inc.
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11018  row: e  column: 18
            High quality sequence stop: 700.
FEATURES             Location/Qualifiers
     source          1. .852
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4995065"
                     /cell_line="MGC36"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_10"
                     /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.5 kb. Library prepared by Life
                     Technologies."
BASE COUNT      238 a    196 c    286 g    132 t
ORIGIN

Alignment Scores:
Pred. No.:                5.2e-57        Length:          852
Score:                    1120.00        Matches:         226
Percent Similarity:       98.70%         Conservative:    1
Best Local Similarity:    98.26%         Mismatches:      3
Query Match:              58.64%         Indels:          3
DB:                       12             Gaps:            0

US-09-972-758A-2 (1-359) x BI091005 (1-852)

Qy        130 LeuGlyAlaProAlaAlaGlyGlyGluGluGluTrpGlyGlnGlnGlnArgGlnLeuGly 149
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 GTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAGTGGGGACAGCAGCAGAGACAGCTGGGG 64

Qy        150 LysLysLysHisArgArgArgProSerLysLysLysArgHisTrpLysProTyrTyrLys 169
              ||||||||||||||||||||||||||||||||||||||||    ||||||||||||||||||
Db         65 AAGAAAAAACATAGGAGACGCCCGTCCAAGAAGAAGCGGCA-TGGAAACCGTACTACAAG 123

Qy        170 LeuThrTrpGluGluLysLysLysPheAspGluLysGlnSerLeuArgAlaSerArgIle 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 CTGACCTGGGAAGAGAAGAAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATC 183

Qy        190 ArgAlaGluMetPheAlaLysGlyGlnProValAlaProTyrAsnThrThrGlnPheLeu 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 CGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTC 243

Qy        210 MetAspAspHisAspGlnGluGluProAspLeuLysThrGlyLeuTyrSerLysArgAla 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 ATGGATGATCACGACCAGGAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCC 303

Qy        230 AlaAlaLysSerAspAspThrSerAspAspAspPheMetGluGluGlyGlyGluGluAsp 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 GCCGCCAAATCCGACGACACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGAT 363

Qy        250 GlyGlySerAspGlyMetGlyGlyAspGlySerGluPheLeuGlnArgAspPheSerGlu 269
              ||||||||||||||||||||||||||||||||||||||||    ||||||||||||||||||
Db        364 GGGGGCAGCGATGGGATGGGAGGGGACGGCAGCGAGTTCTG-CAGCGGGACTTCTCGGAG 422

Qy        270 ThrTyrGluArgTyrHisThrGluSerLeuGlnAsnMetSerLysGlnGluLeuIleLys 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 ACGTACGAGCGGTACCACACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAG 482

Qy        290 GluTyrLeuGluLeuGluLysCysLeuSerArgMetGluAspGluAsnAsnArgLeuArg 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 GAGTACCTGGAACTGGAGAAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGG 542

Qy        310 LeuGluSerLysArgLeuGlyGlyAspAspAlaArgValArgGluLeuGluLeuGluLeu 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 CTGGAGAGCAAGCGGCTGGGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTG 602

Qy        330 AspArgLeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnGluLeuHisArgGlnGln 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||    ||||||
Db        603 GACCGGCTGCGCGCCGAGAACCTCCAGCTGCTGACCGAGAACGAACTGCACGG-CAGCAG 661

Qy        350 GluArgAlaProLeuSerLysPheGlyAsp 359
              |||||||||||||||||||||||||||||
Db        662 GAGCGAGCGCCGCTTTCCAAGTTTGGAGAC 691
```

```
RESULT 12
AL578242/c
LOCUS       AL578242               1078 bp    mRNA    linear   EST 01-JUN-2003
DEFINITION  AL578242 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
            cDNA clone CS0DK002YF07 3-PRIME, mRNA sequence.
ACCESSION   AL578242
VERSION     AL578242.2  GI:31316460
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1078)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 16, 2001 this sequence version replaced gi:12942132.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 8422.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DK002CC04NP1&cluster=8422.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DK002CC04NP1.
FEATURES             Location/Qualifiers
     source          1. .1078
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DK002YF07"
                     /cell_type="HELA CELLS COT 25-NORMALIZED"
                     /cell_line="HELA"
                     /clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      228 a    310 c    232 g    290 t     18 others
ORIGIN

Alignment Scores:
Pred. No.:              6.31e-57        Length:         1078
Score:                  1120.00         Matches:        219
Percent Similarity:     95.63%          Conservative:   0
Best Local Similarity:  95.63%          Mismatches:     10
Query Match:            58.64%          Indels:         1
DB:                     9               Gaps:           0

US-09-972-758A-2 (1-359) x AL578242 (1-1078)

Qy        131 GlyAlaProAlaAlaGlyGlyGluGluGluTrpGlyGlnGlnGlnArgGlnLeuGlyLys 150
              ||||||||||||   |||   ||||||||||||   |||       ||||||||||||||||
Db       1060 GGGGCTCCTGCCSCAGGGGSCGAAGAGGAGTGGGAACAGCWGYAGAGACAGCTGGGGAAG 1001

Qy        151 LysLysHisArgArgArgProSerLysLysLysArgHisTrpLysProTyrTyrLysLeu 170
              |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1000 AAAAWACATAGGAGACGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTACTACAAGCTG 941

Qy        171 ThrTrpGluGluLysLysLysPheAspGluLysGlnSerLeuArgAlaSerArgIleArg 190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        940 ACCTGGGAAGAGAAGAAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGA 881

Qy        191 AlaGluMetPheAlaLysGlyGlnProValAlaProTyrAsnThrThrGlnPheLeuMet 210
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        880 GCCGAGATGTTCGCCAAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTCATG 821

Qy        211 AspAspHisAspGlnGluGluProAspLeuLysThrGlyLeuTyrSerLysArgAlaAla 230
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        820 GATGATCACGACCAGGAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCC 761

Qy        231 AlaLysSerAspAspThrSerAspAspAspPheMetGluGluGlyGlyGluGluAspGly 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 GCCAAATCCGACGACACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGG 701

Qy        251 GlySerAspGlyMetGlyGlyAspGlySerGluPheLeuGlnArgAspPheSerGluThr 270
              ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db        700 GGCAGCGATGGGATGGGAGGRNACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACG 641

Qy        271 TyrGluArgTyrHisThrGluSerLeuGlnAsnMetSerLysGlnGluLeuIleLysGlu 290
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 TACGAGCGGTACCACACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAG 581

Qy        291 TyrLeuGluLeuGluLysCysLeuSerArgMetGluAspGluAsnAsnArgLeuArgLeu 310
              ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db        580 TACCTGGAACTGGAGAAGTGCYTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTG 521

Qy        311 GluSerLysArgLeuGlyGlyAspAspAlaArgValArgGluLeuGluLeuGluLeuAsp 330
              ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db        520 GAGAGCAAGCGGCTGGGTGGCGACTACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGAC 461

Qy        331 ArgLeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnGluLeuHisArgGlnGlnGlu 350
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 CGCTG-CGCGCCGAGAACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAG 402

Qy        351 ArgAlaProLeuSerLysPheGlyAsp 359
              |||||||||||||||||||||||||||
Db        401 CGAGCGCCGCTTTCCAAGTTTGGAGAC 375

RESULT 13
BG291150
LOCUS       BG291150               871 bp    mRNA    linear   EST 21-FEB-2001
DEFINITION  602387182F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4515812 5',
            mRNA sequence.
ACCESSION   BG291150
VERSION     BG291150.1  GI:13048811
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 871)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/..
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10406  row: d  column: 21
            High quality sequence stop: 730.
FEATURES             Location/Qualifiers
     source          1. .871
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4515812"
                     /tissue_type="transitional cell papilloma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_93"
                     /note="Organ: bladder; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.7 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      210 a    230 c    324 g    107 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:                1.68e-56        Length:          871
Score:                    1111.50         Matches:         240
Percent Similarity:       90.07%          Conservative:    5
Best Local Similarity:    88.24%          Mismatches:      15
Query Match:              58.19%          Indels:          15
DB:                       10              Gaps:            2

US-09-972-758A-2 (1-359) x BG291150 (1-871)

Qy          60 GlyProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCys 79
               |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          11 GGGGCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGT 70

Qy          80 ProGluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAla 99
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          71 CCAGAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCT 130

Qy         100 GlyGlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAla 119
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         131 GGCGGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCC 190

Qy         120 GlnProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGlu 139
               ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db         191 CAGCCTTGTCATGACTCCGAGGCCAGTAAGTG-GGGGCTCCTGCCGCAGGGGGCGAAGAG 249

Qy         140 GluTrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLys 159
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         250 GAGTGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAG 309

Qy         160 LysLysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAsp 179
               |||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db         310 AAGAAGCGGCA-TGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGAC 368

Qy         180 GluLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnPro 199
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         369 GAGAAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCG 428

Qy         200 ValAlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAsp 219
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         429 GTCGCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGAT 488

Qy         220 LeuLysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAsp 239
               ||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db         489 CTCAAAACCGGCCTGTACTCCAAGCGGGCGC--GCCAAATCCGACGACACCAGCGATGAC 546

Qy         240 AspPheMetGluGluGlyGlyGluGluAspGlyGlySerAsp-GlyMetGlyGlyAspGl 259
               ||||||||||||||||||||||||||||||||||||||||::: ||||||||||||||||
Db         547 GACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGAATGGGATGGGAGGGGACGG 606

Qy         259 ySerGluPheLeuGlnArgAspPheSerGluThrTyrGlu-ArgTyrHisThrGluSerL 279
               |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db         607 CAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCCGGTACCACACGGAGAGCC 666

Qy         279 euGln-AsnMetSerLysGln-GluLeuIleLys-GluTyrLeuGluLeu-GluLys--- 296
               ||||| :::||||||||||||| |||||||||| |||||||||||||| ||||||
Db         667 TGCAGAAGCATGAGCAAGCAGGGAGCTCATCAAGGGAGTACCTGGAACTGGGAGAAGGTG 726

Qy         297 CysLeuSerArgMetGluAspGluAsnAsnArgLeuArgLeuGlu-------------- 311
                  |||             |||||||||:::|||      :::|||
Db         727 GCTCTCGGGGCCTGGAGGGACGAGAACCACCGGTGCGGTGTGGAGCCACCGGTGGTGGCG 786

Qy         312 SerLysArgLeuGlyGlyAspAsp 319
                   |||||||||      :::
Db         787 AGACCCCGTTTGGGAGCGGGGGAG 810

RESULT 14
BM474968
LOCUS       BM474968               1294 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6476735 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5562828
            5', mRNA sequence.
ACCESSION   BM474968
VERSION     BM474968.1  GI:18524010
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1294)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12292  row: n  column: 13
            High quality sequence stop: 507.
FEATURES             Location/Qualifiers
     source          1. .1294
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5562828"
                     /tissue_type="duodenal adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_88"
                     /note="Organ: small intestine; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
                     oligo-dT primed. Average insert size 1.767 kb. Library
                     enriched for full-length clones and constructed by Life
                     Technologies. Note: this is a NIH_MGC Library."
BASE COUNT      263 a    380 c    407 g    241 t      3 others
ORIGIN

Alignment Scores:
Pred. No.:                3.79e-54        Length:          1294
Score:                    1074.00         Matches:         214
Percent Similarity:       89.80%          Conservative:    6
Best Local Similarity:    87.35%          Mismatches:      14
Query Match:              56.23%          Indels:          11
DB:                       12              Gaps:            3

US-09-972-758A-2 (1-359) x BM474968 (1-1294)

Qy          31 GluArgProProGlyAlaGluGluArgValProGluGluAspSerArgTrpGlnSerArg 50
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           3 GAGCGCCCCCCAGGCGCGGAGGAGCGGGTGCCCGAGGAGGACAGTAGGTGGCAATCGAGA 62

Qy          51 AlaProProGlnLeuGlyGlyArgProGlyProGluGlyGluGlySerLeuGluSerGln 70
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          63 GCGTTCCCCCAGTTGGGTGGCCGTCCGGGGCCGGAGGGGGAAGGGAGCCTGGAATCCCAA 122

Qy          71 ProProProLeuGlnThrGlnAlaCysProGluSerSerCysLeuArgGluGlyGluLys 90
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         123 CCACCTCCCTTGCAGACCCAGGCCTGTCCAGAATCTAGCTGCCTGAGAGAGGGCGAGAAG 182

Qy          91 GlyGlnAsnGlyAspAspSerSerAlaGlyGlyAspProProProAlaGluValGlu 110
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         183 GGCCAGAATGGGGACGACTCGTCCGCTGGCGGCGACTTCCCGCCGCCGGCAGAAGTGGAA 242

Qy         111 ProThrProGluAlaGluLeuLeuAlaGlnProCysHisAspSerGluAlaSerLysLeu 130
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         243 CCGACGCCCGAGGCCGAGCTGCTCGCCCAGCCTTGTCATGACTCCGAGGCCAGTAAGTTG 302

Qy         131 GlyAlaProAlaAlaGlyGlyGluGluGluTrpGlyGlnGlnGlnArgGlnLeuGlyLys 150
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         303 GGGGCTCCTGCCGCAGGGGGCGAAGAGGAGTGGGGACAGCAGCAGAGACAGCTGGGGAAG 362

Qy         151 LysLysHisArgArgArgProSerLysLysLysArgHisTrpLysProTyrTyrLysLeu 170
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 AAAAAACATAGGAGACGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTACTACAAGCTG 422

Qy        171 ThrTrpGluGluLysLysLysPheAspGluLysGlnSerLeuArgAlaSerArgIleArg 190
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 ACCTGGGAAGAGAAGAAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGA 482

Qy        191 AlaGluMetPheAlaLysGlyGlnProValAlaProTyrAsnThrThrGln-PheLeuMe 210
                |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        483 GCCGAGATGTTCGCCAAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTTCCTCAT 542

Qy        210 tAspAspHisAspGlnGluGluProAspLeuLysThrGlyLeuTyrSerLysArg-AlaA 230
                |||||||||||||||||||||||||||||||    |||||||||||||||:::||| ||||
Db        543 GGATGATCACGACCAGGAGGAGCCGGATCTCCNAACCGGCCTGTACTCCCAGCGGGGCCG 602

Qy        230 la-AlaLysSerAspAspThrSerAspAsp---------AspPheMetGluGluGlyGly 246
                || |||||||||     |||      |||                :::   |||
Db        603 CCCGCCAAATCCCGGACGACACCAGCCGATTGACCACTTCATGGGAANCAAGGGGGGTGG 662

Qy        247 GluGluAspGlyGly---SerAspGlyMetGlyGlyAspGlySerGlu------------ 261
                ::::::|||||||||   :::|||||||      |||   |||:::
Db        663 AAGAAGGATGGGGGGCCAGCCGATGGGGAATGGGGAAGGGGGAACGGCCGGCCCAATTTT 722

Qy        262 PheLeuGln 264
                ||||||||
Db        723 TTCTTGCAG 731

RESULT 15
CA430432
LOCUS       CA430432                683 bp    mRNA    linear   EST 07-NOV-2002
DEFINITION  UI-H-DF0-bev-f-23-0-UI.s1 NCI_CGAP_DF0 Homo sapiens cDNA clone
            UI-H-DF0-bev-f-23-0-UI 3', mRNA sequence.
ACCESSION   CA430432
VERSION     CA430432.1  GI:24793158
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 683)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jose Mercuende
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Clone distribution information can be obtained
            from Dr. M. Bento Soares, bento-soares@uiowa.edu
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .683
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-H-DF0-bev-f-23-0-UI"
                     /tissue_type="Subchondral Bone"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI_CGAP_DF0"
                     /note="Organ: Bone; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     NCI_CGAP_DF0 is a cDNA library containing the following
                     tissue(s): Subchondral Bone. The library was constructed
                     according to Bonaldo, Lennon and Soares, Genome Research,
                     6:791-806, 1996. First strand cDNA synthesis was primed
                     with an oligo-dT primer containing a Not I site. Double
                     stranded cDNA was ligated to an EcoR I adaptor, digested
                     with Not I, and cloned directionally into pT7T3-Pac
                     vector. The oligonucleotide used to prime the synthesis of
                     first-strand cDNA contains a library tag sequence that is
                     located between the Not I site and the (dT)18 tail. The
                     sequence tag for this library is GTTAAGCGTC.
                     TAG_LIB=UI-H-DF0
                     TAG_TISSUE=subchondral bone
                     TAG_SEQ=GTTAAGCGTC"
BASE COUNT      170 a    182 c    211 g    120 t
ORIGIN

Alignment Scores:
Pred. No.:              1.72e-53        Length:         683
Score:                  1059.00         Matches:        195
Percent Similarity:     98.99%          Conservative:   1
Best Local Similarity:  98.48%          Mismatches:     2
Query Match:            55.45%          Indels:         0
DB:                     14              Gaps:           0

US-09-972-758A-2 (1-359) x CA430432 (1-683)

Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         90 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 149

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        150 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 209

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaProProGlnLeuGlyGlyArgProGly 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        210 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 269

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        270 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 329

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        330 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 389

Qy        101 GlyAspProProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        390 GGCGACTTCCCGCCGCCGGCAGAAGTGGAGCCGACGCCCGAGGCCGAGCTGCTCGCCCAG 449

Qy        121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        450 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 509

Qy        141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        510 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 569

Qy        161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        570 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 629

Qy        181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGln 198
                ||||||||||||   :::   ||||||||||||||||||||||||||||||
Db        630 AAACAGAGCCTTCTGAGCTTCAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAG 683


Search completed: February  5, 2004, 15:50:21
Job time : 2664 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: February 5, 2004, 12:03:48 ; Search time 81 Seconds
(without alignments)
703.492 Million cell updates/sec

Title: US-09-972-758A-2
Perfect score: 1910
Sequence: 1 MAEPFLSEYQHQPQTSNCTG.....LTENELHRQQERAPLSKFGD 359

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1910	100.0	359	19	AAW85455	Secreted protein e
2	1910	100.0	359	23	ABP61799	Human polypeptide
3	1910	100.0	359	23	ABB76495	Human oestrogen do
4	1903	99.6	359	22	AAB95465	Human protein sequ
5	545.5	28.6	286	23	ABP43855	Mouse putative pro
6	308.5	16.2	152	22	AAM25726	Human protein sequ
7	237	12.4	134	22	ABG15306	Novel human diagno
8	209	10.9	349	22	ABB59958	Drosophila melanog
9	155.5	8.1	281	22	ABG15303	Novel human diagno
10	147	7.7	954	22	AAU14615	Novel bone marrow
11	143	7.5	288	22	ABG20362	Novel human diagno
12	141.5	7.4	1192	22	ABB59642	Drosophila melanog
13	138	7.2	557	19	AAY20666	Human neurofilamen
14	135.5	7.1	530	21	AAY94461	Human p75 protein.
15	135.5	7.1	530	23	AAU74503	Human nuclear tran
16	135.5	7.1	1749	22	ABG00839	Novel human diagno
17	135	7.1	554	23	ABB04721	Human PP2464 prote
18	135	7.1	555	22	AAM93869	Human polypeptide,
19	134.5	7.0	510	22	ABB11764	Human LDL binding
20	134.5	7.0	510	22	AAM79741	Human protein SEQ
21	134.5	7.0	546	22	AAB82808	Human low density
22	134.5	7.0	639	24	ABR41186	Human DITHP extrac
23	134.5	7.0	639	24	ABR41210	Human DITHP extrac
24	133	7.0	538	21	AAB10043	MMLV gag protein.
25	133	7.0	648	20	AAY17946	MoMLV gag gene pro
26	133	7.0	1737	21	AAB10044	MMLV gag-pol prote
27	132.5	6.9	374	22	AAG64026	Human polypeptide
28	132.5	6.9	530	20	AAW97775	Human lens epithel
29	132.5	6.9	530	23	AAU74501	Human nuclear anti
30	132.5	6.9	530	23	AAU74502	Human lens epithel
31	132	6.9	528	22	AAM78918	Human protein SEQ
32	131	6.9	932	22	ABB65256	Drosophila melanog
33	131	6.9	1424	22	AAM39253	Human polypeptide
34	131	6.9	1464	22	AAM41039	Human polypeptide
35	131	6.9	1898	20	AAY30795	A human trichohyal
36	130.5	6.8	819	22	AAB94316	Human protein sequ
37	130	6.8	611	20	AAY29039	T. gondii immunoge
38	130	6.8	611	22	AAU25510	T. gondii immunoge
39	129.5	6.8	1743	22	ABG10928	Novel human diagno
40	128	6.7	670	23	ABP62902	Human polypeptide
41	127.5	6.7	984	22	AAO12986	Human polypeptide
42	127	6.6	654	22	ABB63266	Drosophila melanog
43	126.5	6.6	470	24	AAE33670	Human structural a
44	126.5	6.6	661	22	ABB61881	Drosophila melanog
45	126.5	6.6	710	22	ABG20363	Novel human diagno

ALIGNMENTS

RESULT 1
AAW85455
ID AAW85455 standard; Protein; 359 AA.
XX
AC AAW85455;
XX
DT 25-FEB-1999 (first entry)
XX
DE Secreted protein encoded by clone bp783_3.
XX
KW Secreted protein; nutritional activity; immune stimulating; vaccine;
KW suppressing activity; haematopoiesis regulating activity;
KW tissue growth activity; activin; inhibin activity; chemotactaxis;
KW chemokinetic activity; haemostasis; thrombolytic activity; receptor;
KW ligand; anti-inflammatory; cadherin; tumour invasion suppressor;
KW tumour inhibition; gene therapy.
XX
OS Homo sapiens.
XX
PN WO9842739-A2.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US05653.
XX
PR 19-MAR-1998; 98US-0044466.
PR 21-MAR-1997; 97US-0822167.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;

XX
DR WPI; 1998-609890/51.
DR N-PSDB; AAV82778.
XX
PT New polynucleotides encoding secreted human proteins - derived from
PT human foetal brain, adult brain, foetal kidney, placenta or adult
PT pineal gland cDNA libraries.
XX
PS Claim 1; Page 67-68; 113pp; English.
XX
CC The present sequence represents a secreted protein. The polynucleotide
CC and secreted protein are predicted to have biological activities which
CC would make them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is given.
CC Suggested activities include nutritional activity, immune stimulating
CC (e.g. as vaccines) or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC invasion suppressor activity, and tumour inhibition activity (no data is
CC given in the specification to support these activities). The
CC polynucleotide is also stated to be useful for gene therapy.
XX
SQ Sequence 359 AA;

Query Match 100.0%; Score 1910; DB 19; Length 359;
Best Local Similarity 100.0%; Pred. No. 6.5e-153;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAEPFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAEPFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPG 60

Qy         61 PEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELLAQ 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELLAQ 120

Qy        121 PCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDE 180

Qy        181 KQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDD 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDD 240

Qy        241 FMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCLSR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 FMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCLSR 300

Qy        301 MEDENNRLRLESKRLGGDDARVRELELELDRLRAENLQLLTENELHRQQERAPLSKFGD 359
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 MEDENNRLRLESKRLGGDDARVRELELELDRLRAENLQLLTENELHRQQERAPLSKFGD 359
```

RESULT 2
ABP61799
ID ABP61799 standard; Protein; 359 AA.
XX
AC ABP61799;
XX
DT 04-OCT-2002 (first entry)
XX
DE Human polypeptide SEQ ID NO 153.
XX
KW Human; cytostatic; antirheumatic; antiarthritic; vulnerary; analgesic;
KW antiinflammatory; antibacterial; immunosuppressive; antiparkinsonian;
KW neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;
KW antiulcer; fungicide; antidiabetic; antiasthmatic; antiallergic;
KW immunostimulant; antiparasitic; secreted protein; transmembrane protein;
KW cytokine; cell proliferation; cell differentiation; autoimmune disease;
KW stem cell; growth factor; nervous system disease; neuropathy;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW osteoporosis; severe combined immunodeficiency; SCID; infection;
KW multiple sclerosis; rheumatoid arthritis; gene therapy.
XX
OS Homo sapiens.
XX
PN US2002065394-A1.
XX
PD 30-MAY-2002.
XX
PF 22-DEC-2000; 2000US-0745763.
XX
PR 18-MAR-1998; 98US-0040963.
XX
PA (JACO/) JACOBS K.
PA (MCCO/) MCCOY J M.
PA (LAVA/) LAVALLIE E R.
PA (COLL/) COLLINS-RACIE L A.
PA (EVAN/) EVANS C.
PA (MERB/) MERBERG D.
PA (TREA/) TREACY M.
PA (SPAU/) SPAULDING V.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Spaulding V;
XX
DR WPI; 2002-582343/62.
DR N-PSDB; ABQ92015.
XX
PT Novel secreted or transmembrane protein and polynucleotide encoding the
PT protein, useful for diagnosis and treatment of neurological disorders,
PT cancer, autoimmune diseases, bone disorders and lung or liver fibrosis
PT -
XX
PS Claim 44; Page 112-113; 284pp; English.
XX
CC The invention relates to human secreted or transmembrane protein (I),
CC their fragments and is encoded by specific complementary deoxyribonucleic
CC acid (cDNA) inserts (II), where the protein is substantially free from
CC other mammalian proteins. (I) are useful for preventing, treating or
CC ameliorating a medical condition, especially immunological treatment or
CC prevention of tumours. (I) exhibits activity relating to angiogenesis,
CC cytokine, cell proliferation, cell differentiation, antiinflammatory,
CC stem cell growth factor activity and activin or inhibin-related
CC activities. (I) can be used to manipulate stem cells in culture to give
CC rise to neuroepithelial cells that can be used to augment or replace
CC cells damaged by illness, autoimmune disease, accidental damage or
CC genetic disorders. (I) induces the proliferation of neural cells and
CC regeneration of nerve and brain tissue and is useful for the treatment of
CC central and peripheral nervous system diseases and neuropathies, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis. (I) is involved in chemotactic or chemokinetic
CC activity, regulation of haematopoiesis and is useful for treating myeloid
CC or lymphoid cell disorders, platelet disorders such as thrombocytopaenia
CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve
CC tissue growth and in tissue repair, healing of burns, incisions, ulcers,
CC for treating osteoporosis, osteoarthritis, bone degenerative disorders or
CC periodontal disease. (I) is also useful for gut protection or
CC regeneration and treatment of lung or liver fibrosis, reperfusion injury
CC in various tissues, various immune deficiencies and disorders including
CC severe combined immunodeficiency (SCID), bacterial or fungal infections,
CC autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis,
CC diabetes mellitus, myasthenia gravis, allergic reactions and conditions,
CC such as asthma or other respiratory problems. (II) is useful to express
CC recombinant protein, as markers for tissues in which the corresponding
CC protein is preferentially expressed and in gene therapy. The present
CC sequence is that of a polypeptide of the invention.
XX
SQ Sequence 359 AA;

Query Match 100.0%; Score 1910; DB 23; Length 359;
Best Local Similarity 100.0%; Pred. No. 6.5e-153;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAEPFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPG 60
```

```
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAEPFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPG 60

Qy         61 PEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELLAQ 120
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELLAQ 120

Qy        121 PCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDE 180
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDE 180

Qy        181 KQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDD 240
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDD 240

Qy        241 FMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCLSR 300
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 FMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCLSR 300

Qy        301 MEDENNRLRLESKRLGGDDARVRELELELDRLRAENLQLLTENELHRQQERAPLSKFGD 359
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 MEDENNRLRLESKRLGGDDARVRELELELDRLRAENLQLLTENELHRQQERAPLSKFGD 359
```

RESULT 3
ABB76495
ID ABB76495 standard; Protein; 359 AA.
XX
AC ABB76495;
XX
DT 10-SEP-2002 (first entry)
XX
DE Human oestrogen downregulated gene EDG1 protein.
XX
KW EDG1; oestrogen downregulated gene; tumour suppressor; human;
KW breast cancer; prostate cancer; testicular cancer; ovarian cancer;
KW uterine cancer; colon cancer; gene therapy.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT Region   150..177
FT   /note= "nuclear localisation signal"
FT Misc-difference 154
FT   /note= "encoded by AAG"
FT Misc-difference 171
FT   /note= "encoded by AAC"
XX
PN WO200228879-A1.
XX
PD 11-APR-2002.
XX
PF 05-OCT-2001; 2001WO-US31300.
XX
PR 05-OCT-2000; 2000US-238187P.
XX
PA (UYCA-) UNIV CASE WESTERN RESERVE.
XX
PI Montano M, Wittman B;
XX
DR WPI; 2002-519107/55.
DR N-PSDB; ABN84013.
XX
PT Polynucleotides encoding Estrogen Down-Regulated Gene 1 proteins,
PT useful for the prevention, diagnosis and treatment of e.g. breast
PT cancer, testicular cancer, prostate cancer, uterine cancer, cervical
PT cancer, ovarian cancer and colon cancer -
XX
PS Claim 7; Fig 1A-B; 52pp; English.
XX
CC The present sequence is the protein sequence of human EDG1, a
CC 40 kDa protein encoded by the tumour suppressor gene designated
CC human oestrogen downregulated gene 1 (see ABN84013). EDG1 mRNA
CC expression is prevalent in normal mammary epithelial cells and in
CC other human hormone-responsive tissues such as the ovary, prostate
CC and testis. Expression is low in breast cancer epithelial cells.
CC Oestradiol, which induces breast cancer cell growth, has an
CC inhibitory effect on EDG1 mRNA expression in breast cancer cells.
CC Hexamethylene bis-acetamide, an inducer of differentiation and
CC apoptosis, upregulates EDG1 mRNA expression in breast cancer cells.
CC The invention provides EDG1 polynucleotides and polypeptides. In
CC a claimed method of detecting cancerous cells, a test sample from
CC an individual suspected of having, or known to have breast,
CC testicular, prostate, uterine, cervical, ovarian or colon cancer is
CC contacted with an anti-EDG1 antibody. A decrease in the level
CC or antigen-antibody complex compared to the level of a control
CC sample indicates cancerous cells. A claimed method for decreasing
CC the proliferation of breast, prostate, testicular, ovarian, uterine,
CC cervical or colon cancer cells involves increasing EDG1 protein
CC activity in the cells, either by contacting the cells with EDG1
CC protein, its fragment or functional equivalent, or with a nucleic
CC acid encoding EDG1 protein, its fragment or functional equivalent.
XX
SQ Sequence 359 AA;

```
  Query Match             100.0%;  Score 1910;  DB 23;  Length 359;
  Best Local Similarity   100.0%;  Pred. No. 6.5e-153;
  Matches  359;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAEPFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPG 60
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAEPFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPG 60

Qy         61 PEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELLAQ 120
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELLAQ 120

Qy        121 PCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDE 180
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDE 180

Qy        181 KQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDD 240
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDD 240

Qy        241 FMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCLSR 300
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 FMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCLSR 300

Qy        301 MEDENNRLRLESKRLGGDDARVRELELELDRLRAENLQLLTENELHRQQERAPLSKFGD 359
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 MEDENNRLRLESKRLGGDDARVRELELELDRLRAENLQLLTENELHRQQERAPLSKFGD 359
```

RESULT 4
AAB95465
ID AAB95465 standard; Protein; 359 AA.
XX
AC AAB95465;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:17953.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.

```
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 17953; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   359 AA;

  Query Match             99.6%;  Score 1903;  DB 22;  Length 359;
  Best Local Similarity   99.7%;  Pred. No. 2.6e-152;
  Matches  358;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MAEPFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAEPFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPG 60

Qy         61 PEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELLAQ 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELLAQ 120

Qy        121 PCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDE 180

Qy        181 KQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDD 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDD 240

Qy        241 FMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCLSR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 FMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCLSR 300

Qy        301 MEDENNRLRLESKRLGGDDARVRELELELDRLRAENLQLLTENELHRQQERAPLSKFGD 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        301 MEDENNRLRLESKRLGGDDARVRELELELDRLRAENLQLLTENELHRQQERAPPSKFGD 359


RESULT 5
ABP43855
ID   ABP43855 standard; Protein; 286 AA.
XX
AC   ABP43855;
XX
DT   26-FEB-2003  (first entry)
XX
DE   Mouse putative protein #15.
XX
KW   Neuroprotective; immunomodulator; cancer;
KW   cytostatic; anti-inflammatory; gene therapy; nutritional supplement;
KW   wound; burn; ulcer; Alzheimer's disease; Huntington's disease;
KW   amyotrophic lateral sclerosis; autoimmune disorder; inflammation;
KW   vulnerary.
XX
OS   Mus musculus.
XX
PN   WO200231111-A2.
XX
PD   18-APR-2002.
XX
PF   11-OCT-2001; 2001WO-US27760.
XX
PR   12-OCT-2000; 2000US-0687527.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Zhou P,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Wehrman T,  Drmanac RT;
XX
DR   WPI; 2002-426278/45.
DR   N-PSDB; ABQ61099.
XX
PT   New polypeptides and their encoded proteins, useful as nutritional
PT   sources or supplements, or in gene therapy, particularly for treating
PT   wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
PT   inflammation  -
XX
PS   Claim 20; SEQ ID # 758; 357pp + sequence listing; English.
XX
CC   The invention relates to 446 newly isolated polynucleotide sequences.
CC   The activity of polynucleotides of the invention may be described as,
CC   vulnerary, neuroprotective, immunomodulator, cytostatic and
CC   anti-inflammatory. Compositions comprising nucleic acids of the invention
CC   are useful for treating a mammalian subject, or as nutritional sources or
CC   supplements. These are useful in gene therapy, particularly for treating
CC   wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
CC   amyotrophic lateral sclerosis, autoimmune disorders, cancer or
CC   inflammation. The nucleic acids and polypeptides are also useful in
CC   diagnostic and research methods. The sequences given in records
CC   ABP43544-ABP43989 represent polypeptides encoded by polynucleotides of
CC   the invention.
CC   NOTE: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   286 AA;

  Query Match             28.6%;  Score 545.5;  DB 23;  Length 286;
  Best Local Similarity   43.8%;  Pred. No. 8.4e-38;
  Matches  128;  Conservative   36;  Mismatches   93;  Indels   35;  Gaps    6;

Qy         73 PLQTQACPESSCLREGEKGQNGDDS--------SAGGDFPPPAEVEPTPEAELLAQPC-- 122
              | ||     ||     |  |  |          :||  |    :|   | | ||
Db          5 PNQTACNAESPVALEEAKTSGAPGSPQTPPERHDSGGSLPLTPRMESHSEDEDLAGAVGG 64

Qy        123 --HDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDE 180
                 :| : :  :|     |        | :|||||||||:||||:|| :|:| ||:: ||
Db         65 LGWNSRSPRTQSPGGCSAE------AVLARKKHRRRPSKRKRHWRPYLELSWAEKQQRDE 118

Qy        181 KQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEP--DLKTGLYSKRAAAKSDDTSD 238
              :|| ||||:| ||||||||||||||||||:| | |||  |:  |:    :: :|
Db        119 RQSQRASRVREEMFAKGQPVAPYNTTQFLMNDRDPEEPNLDVPHGISHPGSSGES----- 173
```

```
Qy        239 DDFMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCL 298
                  | | ||| |    || ::||||||||:|||||  |||||:::|||||| |
Db        174 ---------EAGDSDGRGRAHGEFQRKDFSETYERFHTESLQGRSKQELVRDYLELEKRL 224

Qy        299 SRMEDENNRLRLESKRLGGDDAR-VRELELELDRLRAENLQLLTENELHRQQ 349
              |: |:|  ||:       |    | | ||  |: ||| || :|  ||::  ::
Db        225 SQAEEETRRLQQLQACTGQQSCRQVEELAAEVQRLRTENQRLRQENQMWNRE 276


RESULT 6
AAM25726
ID   AAM25726 standard; Protein; 152 AA.
XX
AC   AAM25726;
XX
DT   16-OCT-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:1241.
XX
KW   Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW   antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW   antibacterial; endocrine; cardiant; central nervous system; virucide;
KW   anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW   antiaggregant; haemostatic; vulnerary; antiulcer; osteopathic; eczema;
KW   dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW   neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW   immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW   antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW   cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW   genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW   thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW   allergic rhinitis; diabetes; multiple sclerosis; depression;
KW   Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW   neurological disorder.
XX
OS   Homo sapiens.
XX
PN   WO200153455-A2.
XX
PD   26-JUL-2001.
XX
PF   22-DEC-2000; 2000WO-US35017.
XX
PR   23-DEC-1999;   99US-0471275.
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-457603/49.
DR   N-PSDB; AAH99667.
XX
PT   Isolated human polynucleotides encoding polypeptides, useful for the
PT   treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX
PS   Claim 20; Page 257; 1217pp; English.
XX
CC   AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC   AAM25963. The proteins can have activities based on the tissues and
CC   cells they are expressed in, such as: antiinflammatory; antirheumatic;
CC   antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
CC   central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC   cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;
CC   antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
CC   antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC   antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC   encoding them can be used in gene therapy, antisense therapy and vaccine
CC   production, The proteins and polynucleotides are useful for screening for
CC   agonists or antagonists of a protein and for the treatment and diagnosis
CC   of disorders associated with the activity of a protein e.g. inflammation,
CC   rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC   neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC   infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC   anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC   osteoporosis, severe combined immunodeficiency, eczema, allergic
CC   rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC   Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC   neurological disorders.
XX
SQ   Sequence   152 AA;

  Query Match             16.2%;  Score 308.5;  DB 22;  Length 152;
  Best Local Similarity   46.4%;  Pred. No. 3.8e-18;
  Matches   70;  Conservative   21;  Mismatches   43;  Indels   17;  Gaps    3;

Qy        202 PYNTTQFLMDDHDQEEP--DLKTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSDGMGGDG 259
              |||||||||:| | |||  |:  |:    :: :|              | | ||| |
Db          6 PYNTTQFLMNDRDPEEPNLDVPHGISHPGSSGES-------------EAGDSDGRGRAH 51

Qy        260 SEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCLSRMEDENNRLRLESKRLGGDD 319
              || ::||||||||:||||||  |||||:::|||||| ||: |:|  ||:      |
Db         52 GEFQRKDFSETYERFHTESLQGRSKQELVRDYLELEKRLSQAEEETRRLQQLQACTGQQS 111

Qy        320 AR-VRELELELDRLRAENLQLLTENELHRQQ 349
              | | ||  |: ||| || :|  ||::  ::
Db        112 CRQVEELAAEVQRLRTENQRLRQENQMWNRE 142


RESULT 7
ABG15306
ID   ABG15306 standard; Protein; 134 AA.
XX
AC   ABG15306;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #15297.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS79493.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 45665; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
```

```
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   134 AA;

  Query Match            12.4%;  Score 237;  DB 22;  Length 134;
  Best Local Similarity   40.5%;  Pred. No. 3.5e-12;
  Matches   53;  Conservative   19;  Mismatches   51;  Indels    8;  Gaps    3;

Qy         58 RPGPEGEGSLESQPPPLQTQACP---ESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPE 114
              | | | :||| ::|| :|:   |     :  |    :   :||  |    :|   |
Db          7 RFGTEDDGSLRTRPPVMQSHQWPWRRPRPLVSRGRPQTTPERHDSGGSLPLTPRMESHSE 66

Qy        115 AELLAQPCHDSEASKLGAPAAGGEEEWGQQQRQ--LGKKKHRRRPSKKKRHWKPYYKLTW 172
               | ||: |   ||  :| |   |    |  |||  | :||||||||:||||:|| :|:|
Db         67 DEDLAEGCRWPEAGTVGVP---GPRAQGAAQRQAVLARKKHRRRPSKRKRHWRPYLELSW 123

Qy        173 EEKKKFDEKQS 183
               ||:: ||:||
Db        124 AEKQQRDERQS 134


RESULT 8
ABB59958
ID   ABB59958 standard; Protein; 349 AA.
XX
AC   ABB59958;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 6666.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL04061.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 6666; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   349 AA;

  Query Match            10.9%;  Score 209;  DB 22;  Length 349;
  Best Local Similarity   25.9%;  Pred. No. 2.9e-09;
  Matches   68;  Conservative   43;  Mismatches   84;  Indels   68;  Gaps   10;

Qy         85 LREGEKGQNGD-----DSSAGGDFPPPAEVEPTPEAELLAQPCHDSEASKLGAPAAGGEE 139
              : |  | ::|      ||  ||                          ||  |  | ||
Db          1 MAEAVKNESGSQQRPLDSGGGGG----------------------ASGGGGVAVGGGS 36

Qy        140 EWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWE----EKKKFDEKQSLRASRIRAEMFA 195
                 ::: : |||  : :| | | |: | :||          : :::|: |   :|:
Db         37 GMPKRKHRRGKKS-KMQPKKTKNHY-PQWKLDMSTGAGATLEGNQRQNSRTKLVRSRSL- 93

Qy        196 KGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSDGM 255
                : |||| :|||::|  |                ||: |: |   | : :
Db         94 ----LVPYNTNRFLMEEHMSE-------------LHKDDSDDNCF---GSQTE------ 126

Qy        256 GGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCLSRMEDENNRLRLESKRL 315
               |   || ::||: |||   | |: |||||||:| :::|   |: :: :        :|
Db        127 --DQVLFLSKEFSDVYERARLERLETMSKQELIQECMQIEDRYSKAQNISKEF---GAKL 181

Qy        316 GGDDARVRELELELDRLRAENLQ 338
               | ::|:|  |   ||   |:
Db        182 RAQDDKIRQLSRENQFLRTHLLR 204


RESULT 9
ABG15303
ID   ABG15303 standard; Protein; 281 AA.
XX
AC   ABG15303;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #15294.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS79490.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
```

```
PS   Claim 20; SEQ ID No 45662; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   281 AA;

  Query Match             8.1%;  Score 155.5;  DB 22;  Length 281;
  Best Local Similarity   26.0%;  Pred. No. 7.1e-05;
  Matches   60;  Conservative   24;  Mismatches   86;  Indels   61;  Gaps   12;

Qy         14 QTSNCTG-AAAVQEELNPERPPGAEERVPEE-------DSRWQSRAFPQLGGRPGP-EGE 64
              : | | |    ::   :|  | |||||||         : || :||  |:||   |  ||
Db         67 RASACEGDPTLLRPSHSPGHPRGAEERVPSTACLQRSWEGRW-NRALQQVGGLKSPWRGE 125

Qy         65 GSLESQPPP-----LQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEV-----EPTPE 114
                 |||       || |||    | |:| :|              | ::      ||
Db        126 YKEPRHPPPNQANYHQTLACPAERCFRQGPQG-------------PGQLRRQQAEPL-S 170

Qy        115 AELLAQPCHDSEASKLGAPAAGGE---------EEWGQQQRQLGKKKHRRRPSKKKR--H 163
              : ||:      |: | |   |                | |   :    :  ::||: :
Db        171 SSLLSHQSPQSQTSLLPXLXARRSLGCGNPYLCPPWSQPLPETPAHQSGKQPSRSRNCSS 230

Qy        164 WKPYYKLT-------WEEKKKFDEKQSLRASRIR--------AEMFAKGQP 199
              ||  |  |       |  : ||    :  ||  |          : |:| |
Db        231 WKXNYSSTACPSHFLWRSRSKFSRASASTASPCRMTVRRTHQRPLKARGTP 281


RESULT 10
AAU14615
ID   AAU14615 standard; Protein; 954 AA.
XX
AC   AAU14615;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Novel bone marrow polypeptide #14.
XX
KW   Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;
KW   haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;
KW   wound healing; nutritional supplement; immune disorder;
KW   severe combined immunodeficiency; SCID.
XX
OS   Homo sapiens.
XX
PN   WO200157187-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US03782.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   20-JUN-2000; 2000US-0598075.
PR   19-JUL-2000; 2000US-0620325.
PR   30-NOV-2000; 2000US-0250683.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Ford JE,  Boyle BJ,  Tang YT,  Liu C,  Asundi V,  Zhou P,  Xue AJ;
PI   Ren F,  Drmanac RT;
XX
DR   WPI; 2001-488875/53.
DR   N-PSDB; AAS22920.
XX
PT   Nucleic acids encoding bone marrow polypeptides, useful in diagnostic
PT   and gene therapy -
XX
PS   Claim 10; Page 215-217; 392pp; English.
XX
CC   AAU14602-AAU14794 represent novel bone marrow polypeptides of the
CC   invention. The proteins and corresponding coding sequences may be used
CC   in the prevention, diagnosis and treatment of diseases associated with
CC   inappropriate bone marrow polypeptide expression. For example, to treat
CC   disorders associated with decreased expression by rectifying mutations
CC   or deletions in a patient's genome that affect the activity of the
CC   polypeptides by expressing inactive proteins or to supplement the
CC   patient's own production of the polypeptide. Additionally, the nucleic
CC   acids may be used to produce the polypeptides, by inserting the nucleic
CC   acids into a host cell and culturing the cell to express the protein.
CC   The nucleic acid and its complementary sequences may also be used as DNA
CC   probes in diagnostic assays to detect and quantitate the presence of
CC   similar nucleic acid sequences in samples, and therefore which patients
CC   may be in need of restorative therapy. The proteins may also be used as
CC   antigens in the production of antibodies against bone marrow proteins
CC   and in assays to identify modulators of their expression and activity.
CC   The anti-bone marrow protein antibodies and antagonists may also be used
CC   to down regulate expression and activity. The antibodies may also be used
CC   as diagnostic agents for detecting the presence of the protein in samples
CC   (e.g. by enzyme linked immunosorbant assay (ELISA)). The proteins
CC   may be used to regulate haematopoiesis activity, and consequently in the
CC   treatment of myeloid or lymph cell disorders; in tissue regeneration,
CC   such as wound healing; as a nutritional supplement;  and in treatment of
CC   immune disorders such as severe combined immunodeficiency (SCID).
XX
SQ   Sequence   954 AA;

  Query Match             7.7%;  Score 147;  DB 22;  Length 954;
  Best Local Similarity   21.6%;  Pred. No. 0.0018;
  Matches   97;  Conservative   65;  Mismatches  146;  Indels  142;  Gaps   19;

Qy         25 QEELNPERPPGAEER-------VPEEDSRWQS--RAF------PQLGGRPGPEGE----- 64
              |:||  ::  | |||       :| |: | ||  |:       ||     ||||:
Db        189 QDELQSKQSKGLEERYHRLSPPLPHEE-RAQSPPRSLATEEEPPQ-----GPEGQPEWKE 242

Qy         65 ---------GSLESQPPPLQTQA-CPESSCLREGEKGQNGDDSSAGGDFPPPAE----VE 110
                      ||  |    : ||  | ::| :  |:    ::   | :     |   |
Db        243 AEELGEDSAASLSLQLSLQREQAPSPPAACEKGKEQHSQAEELGPGQEEAEDPEEKVAVS 302

Qy        111 PTP-------EAELLAQPCHDSEASKLGAPAAGGEEEWGQQQRQLGKKK----------- 152
              |||          | :| |   |||: | |       :   | || | : |
Db        303 PTPPVSPEVRSTEPVAPPEQLSEAA-LKAMEEAVAQVLEQDQRHLLESKQEKMQQLREKL 361

Qy        153 -----------HRRRPSKKKRHWKPYYKLTWEEKKKFDEKQSLRASRIRAEMFAKGQPVA 201
                        |:::       :   |   ||: :  |::| | | :||:: :  |
Db        362 CQEEEEEILRLHQQKEQSLSSLRERLQKAIEEEEARMREEESQRLSWLRAQVQSSTQA-- 419

Qy        202 PYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSD--------DTSDDDFMEEGGEEDGGS- 252
                    | ||   : :  |   |   :|         :  :   :|:  ||   |
Db        420 ----------DEDQIRAEQEASLQKLREELESQQKAERASLEQKNRQMLEQLKEEIEASE 469

Qy        253 ----------------------DGMGGDGSEFLQRDFSETYERY-------HTE---SLQ 280
                                   :|   :    |::: |   ||         | |   |||
Db        470 KSEQAALNAAKEKALQQLREQLEGERKEAVATLEKEHSAELERLCSSLEAKHREVVSSLQ 529

Qy        281 NMSKQELIKEYLELEKCLSRMEDENNRLRLESKRLGGDDARVREL------------ELE 328
                ::   ||  :|:||| ::|   :|:  :|  : | :  :  |             |
Db        530 KKIQEAQQKEEAQLQKCLGQVE---HRVHQKSYHVAGYEHELSSLLREKRQEVEGEHERR 586
```

```
Qy        329 LDRLRAENLQLLT----ENELHRQQERAPL 354
              ||::: |: |::     : |   :::|| |
Db        587 LDKMKEEHQQVMAKAREQYEAEERKQRAEL 616
```

RESULT 11
ABG20362
ID ABG20362 standard; Protein; 288 AA.
XX
AC ABG20362;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #20353.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS84549.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 50721; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 288 AA;

```
  Query Match             7.5%;  Score 143;  DB 22;  Length 288;
  Best Local Similarity   27.5%;  Pred. No. 0.00083;
  Matches   69;  Conservative   24;  Mismatches   64;  Indels   94;  Gaps   14;

Qy         50 RAFPQL---GGRPG---------PEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDS 97
              ||  ||   | |||         |   |:|      |: :|       | ||:||:  :
Db          9 RAGVQLNSQGARPGGSTAGDACRPRATGTL-----ALRLRA-------RSGERGQSALER 56

Qy         98 SAGGDFPPPAEVEPTPEAELLAQPCHDSEASKLGAPAAGGE------EEW-GQQQR---- 146
              |  |  |||                        |||  |||       : | ||  |
Db         57 S--GRAPPPL------------------LRGLGANGAGGARGQGCGDRWRGQGWRPGRF 95

Qy        147 ------QLGKKKHRRRPSKKKRHWKPYYKLTW-----------EEKKKFDEK-------- 181
                   :  ||| ::   ||:   :   |||            | : | |||      .
Db         96 EIKKKLKTAKKKEKKEKKKKQEEEQEKKKLTQIQESQVTSHNKERRSKRDEKLDKKSQAM 155

Qy        182 QSLRASRIR-----AEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDT 236
              : |:| | :     ||: || ||:    |::   || ::|| |       |   : :|  |
Db        156 EELKAEREKRKNRTAELLAKKQPL---KTSEVYSDDEEEEEDD-----KSSEKSDRSSRT 207

Qy        237 SDDDFMEEGGE 247
              |  |  ||  |
Db        208 SSSDEEEEKEE 218
```

RESULT 12
ABB59642
ID ABB59642 standard; Protein; 1192 AA.
XX
AC ABB59642;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 5718.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL03745.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 5718; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1192 AA;

```
  Query Match             7.4%;  Score 141.5;  DB 22;  Length 1192;
  Best Local Similarity   22.3%;  Pred. No. 0.0072;
```

```
  Matches   80;  Conservative   57;  Mismatches  138;  Indels   83;  Gaps   17;

Qy          4 PFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPGPEG 63
              |   |     ::   | :|   |||  |     :::||||      :  |   |    |
Db        252 PLNDELPEDEESPAATESAV--EELEKESEAAMDDQVPEESEIQPEQVQP---GEYQSES 306

Qy         64 EG-------SLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAE 116
              :|         :|:||   : :| ||:    | |               |  |||| || |
Db        307 DGEQAETKPEIEAQP---EVEAQPEAEAQPEAE--------------PQLEVEPQPEVE 348

Qy        117 LLAQPCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKK 176
                :||  :|:      |    : |   |              : : | :|    |  | :
Db        349 --SQPEVESQPEVEAQPEVEPQSEVESQ------------PEAESHSEPE---TQAEVE 390

Qy        177 KFDEKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDT 236
                | :||  :  : |  |: ||              : ||:::    ::: :    ||
Db        391 AQPEVESLPEAESQPE--AESQP--------------EREPEVE----AEKISDNEVDT 429

Qy        237 SDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEK 296
              ::   ||   |  |  ||:       :  : :|  ::  || |::  :|  : | :| |:
Db        430 TEASLMETLVE--GIEDGLTAAMDNLVPEELAEASDKQETE-LESEDQQSPVTEAIE-EQ 485

Qy        297 CLSRMEDENNRLRLESKRLGGDDARVRELELELDRLRAENLQ-LLTENELHRQQERAP 353
                :  :| | :   | :::   |       | | |  :  | : :    | | : | ||
Db        486 AVPEIEQEKEK---EPEQITLAD------ETEQDSAQPSNEEPVEIAPEQHTEAEIAP 534
```

RESULT 13
AAY20666
ID AAY20666 standard; Protein; 557 AA.
XX
AC AAY20666;
XX
DT 22-JUL-1999 (first entry)
XX
DE Human neurofilament-M wild type protein fragment 8.
XX
KW Human; beta-amyloid precursor protein; beta-APP; diagnosis; cancer;
KW frameshift mutation; age-related disease; neurodegenerative disorder;
KW Alzheimer's disease; Down's syndrome; myotonic dystrophy; neuronal;
KW Huntington's disease; multiple sclerosis; alcoholic liver disease;
KW diabetes mellitus type II; microtubule associated protein; Tau; Big Tau;
KW ubiquitin B; apolipoprotein E; MAP2; neurofilament-L; neurofilament-M;
KW neurofilament-F; presenilin I; presenilin II; cellular tumour antigen;
KW glial fibrillary acidic protein; GFAP; p53; semaphorin III; HUPF-1;
KW bcl-2; B-cell leukemia/lymphoma 2 proto-oncogene; HMGP-C; NSP-A;
KW high mobility group protein-C; neuroendocrine specific protein A.
XX
OS Homo sapiens.
XX
PN WO9845322-A2.
XX
PD 15-OCT-1998.
XX
PF 02-APR-1998; 98WO-IB00705.
XX
PR 10-APR-1997; 97US-0043163.
XX
PA (UYUT-) RIJKSUNIV UTRECHT.
PA (ROYA-) ROYAL NETHERLANDS ACAD ARTS & SCI.
PA (UYRO-) UNIV ROTTERDAM ERASMUS.
XX
PI Burbach JPH, Grosveld FG, Van Leeuwen FW;
XX
DR WPI; 1998-609901/51.
DR N-PSDB; AAX75759.
XX
PT Diagnosing disease by detecting frameshift mutations in RNA or
PT corresponding protein mutations - used to diagnose cancer and
PT neurological diseases, particularly Alzheimer's disease, and also
PT for treatment and prevention with specific ribozymes or wild-type
PT RNA
XX
PS Disclosure; Figure 8; 258pp; English.
XX
CC This invention describes a novel method for the diagnosis of a disease
CC caused by, or associated with, an RNA molecule that has a frameshift
CC mutation. The method is used to diagnose age-related diseases, especially
CC cancer and a wide range of neurodegenerative disorders (e.g. Alzheimer's
CC disease, Down's syndrome, myotonic dystrophy, Huntington's disease,
CC multiple sclerosis, alcoholic liver disease, diabetes mellitus type II
CC and many others listed) or susceptibility to these disorders. The method
CC allows a definitive diagnosis of Alzheimer's disease in living patients,
CC at an early stage. It is based on the observation that disease may be
CC caused by mutations in RNA rather than DNA. The invention describes the
CC used of neuronal system RNA molecules, specifically proteins including
CC beta-amyloid precursor protein (beta-APP), the microtubule associated
CC proteins Tau and Big Tau, ubiquitin B, apolipoprotein E, microtubule
CC associated protein 2 (MAP2), neurofilament-L, neurofilament-M,
CC neurofilament-F, presenilin I, presenilin II, glial fibrillary acidic
CC protein (GFAP), the cellular tumour antigen p53, B-cell leukemia/lymphoma
CC 2 (bcl-2) proto-oncogene, semaphorin III, HUPF-1, high mobility group
CC protein-C (HMGP-C) and neuroendocrine specific protein A.
XX
SQ Sequence   557 AA;

```
  Query Match             7.2%;  Score 138;  DB 19;  Length 557;
  Best Local Similarity   22.5%;  Pred. No. 0.0052;
  Matches   80;  Conservative   38;  Mismatches  122;  Indels  116;  Gaps   17;

Qy         31 ERPPGAEERVPEEDSRWQSRAFPQLGGRPGPEGEG--SLESQPPPLQTQACPESSCLREG 88
              || ||       | ||:            || | |  |  |:||||    ||:|   |
Db          4 ERGPGGGAGPRAEKSRFPCGL--------GPRGAGAESPFSRPPPLLPDRCPQSPPPR-- 53

Qy         89 EKGQNGDDSSAGGDFPPPAEVEPTPEAELLAQPCHDSEASKLGAPAAGGEEEWGQQQRQL 148
                           |||    | |                   |  ||
Db         54 ---------------PPPPRGGPWP----------------GTSAAAIN---------- 71

Qy        149 GKKKHRR------RPSKKKRHW-------KPYYKLTWEEKKKFDEKQSLRASRIRAEMFA 195
              | :: |:       ||   |  :          |:   |  :  |       :|  |:: ::
Db         72 GLRRGRQNAVTATRPKASKMSYTLDSLGNPSAYRRVTETRSSFSRVSGSPSSGFRSQSWS 131

Qy        196 KGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKR----AAAKSDDTSDDDFMEEGGEEDGG 251
              :| |                     |  : : |    :|  |   |  || :     :||
Db        132 RGSP-------------STVSSSYKRSMLAPRLAYSSAMLSSAESSLDFSQSSSLLNGG 177

Qy        252 SDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCLSRMEDENNRLRLE 311
              | | |||         | : |:   | || ::  :    |:|    :  :| :|  :  |
Db        178 S-GPGGD------YKLSRSNEK---EQLQGLN--DRFAGYIEK---VHYLEQQNKEIEAE 222

Qy        312 SKRLG---------GD--DARVRELELELDRLRAENLQLL-----TENELHRQQER 351
                : |          ||  |  :|||   |: :  |  |:       | ::|| :||
Db        223 IQALRQKQASHAQLGDAYDQEIRELRATLEMVNHEKAQVQLDSDHLEEDIHRLKER 278
```

RESULT 14
AAY94461
ID AAY94461 standard; Protein; 530 AA.
XX
AC AAY94461;
XX
DT 11-SEP-2000 (first entry)
XX
DE Human p75 protein.
XX
KW p52; p75; transcription coactivator; ASF/SF2; gene therapy;
KW cancer; breast cancer.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT Region   192..325
FT   /label= Highly_charged
XX
PN WO200029578-A1.
XX

```
PD   25-MAY-2000.
XX
PF   10-NOV-1999;   99WO-US26792.
XX
PR   13-NOV-1998;   98US-0108248.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Ge H;
XX
DR   WPI; 2000-387791/33.
DR   N-PSDB; AAA27150.
XX
PT   Novel mRNA transcription coactivators p52 and p75 useful for treating
PT   tumours and cancers comprising a specific amino acid sequence  -
XX
PS   Claim 6; Fig 3; 100pp; English.
XX
CC   The N-terminal and two internal sequences were identified in a 75kDa
CC   polypeptide that co-purified with a general transcription coactivator.
CC   A degenerate oligonucleotide, based on the N-terminal sequence, was used
CC   to screen a human cDNA library derived from an adenocarcinoma of the
CC   cervix. From this the sequences of p52 and p75 were identified. The p52
CC   and to a lesser extent p75 proteins were found to act as general
CC   transcriptional coactivators. In addition p52 was found to enhance
CC   ASF/SF2-mediated pre-mRNA splicing. The p52 and p75 polypeptides can be
CC   used to treat disease caused by defects in transcription or in
CC   ASF/SF2-mediated pre-mRNA splicing. These disease may also be treated by
CC   gene therapy, using p52 and p75 polynucleotides. An example of a
CC   treatable disease is cancer, especially breast cancer. The present
CC   sequence is the human p75 protein, GenBank Accession No.
CC   AAC97946.
XX
SQ   Sequence   530 AA;

  Query Match             7.1%;  Score 135.5;  DB 21;  Length 530;
  Best Local Similarity   21.8%;  Pred. No. 0.008;
  Matches   79;  Conservative   59;  Mismatches  110;  Indels  115;  Gaps   18;

Qy          7 SEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQ----------L 55
              |  |   : || :   |:|:         |  | :||:  : :|  :
Db         93 SSQQAATKQSNASSDVEVEEK---------ETSVSKEDTDHEEKASNEDVTKAVDITTPK 143

Qy         56 GGRPG--PEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEV---E 110
                |  |   : |  :|::   : | |    :     ::|:           |   ||   :
Db        144 AARRGRKRKAEKQVETEEAGVVTTATASVNLKVSPKRGR----------PAATEVKIPK 192

Qy        111 PTPEAELLAQPCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKL 170
              |   ::: ||| ||  ::  :       ||:  ::: | ||     :| |
Db        193 PRGRPKMVKQPC-PSESDII------TEEDKSKKKGQEGK-----QPKK---------- 229

Qy        171 TWEEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTG---LYSK 227
              : || :| |                          :|  ::||| | |   : ||
Db        230 ---QPKKDEEGQK-------------------------EEDKPRKEPDKKEGKKEVESK 260

Qy        228 RA-AAKSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSETYER------YHTESLQ 280
              |   ||:  ||  |  ||| :::|     ||       |:|   : |      :  |:
Db        261 RKNLAKTGVTSTSDSEEEGDDQEGEKKRKGG-------RNFQTAHRRNMLKGQHEKEAAD 313

Qy        281 NMSKQELIKEYLELEKCLSRMEDENNR---LRLESKRLGGDDARVRELELEL-DRLRAEN 336
              |||   | :| |:   : :||  :    ::| ||    |:|:: :  |: : |: :|
Db        314 RKRKQE---EQMETEQ---QNKDEGKKPEVKKVEKKRETSMDSRLQRIHAEIKNSLKIDN 367

Qy        337 LQL 339
              | :
Db        368 LDV 370


RESULT 15
AAU74503
ID   AAU74503 standard; Protein; 530 AA.
XX
AC   AAU74503;
XX
DT   09-APR-2002  (first entry)
XX
DE   Human nuclear transcription co-activator p75.
XX
KW   Human; nuclear antigen; nuclear transcription co-activator; p75;
KW   atopic dermatitis; asthma; interstitial cystitis; protein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 224
FT                   /note= "Encoded by GTA"
XX
PN   WO200190751-A1.
XX
PD   29-NOV-2001.
XX
PF   23-MAY-2001; 2001WO-US16639.
XX
PR   23-MAY-2000; 2000US-207141P.
XX
PA   (SCRI ) SCRIPPS RES INST.
XX
PI   Tan EM,  Ochs RL,  Chan EKL,  Muro Y;
XX
DR   WPI; 2002-122026/16.
DR   N-PSDB; AAS20483.
XX
PT   Screening a patient for atopic dermatitis, comprises determining the
PT   presence of antibodies against nuclear transcription co-activator p75
PT   in the sera of the patient  -
XX
PS   Disclosure; Page -; 22pp; English.
XX
CC   The invention relates to a method of screening a patient for atopic
CC   dermatitis, comprising determining the presence of antibodies against
CC   nuclear transcription co-activator p75, in sera of the patient. The
CC   method is used for diagnosing patients with atopic dermatitis, and for
CC   detecting atopic dermatitis in patients suffering from other conditions
CC   such as asthma or interstitial cystitis. This sequence represents the
CC   human p75 protein.
XX
SQ   Sequence   530 AA;

  Query Match             7.1%;  Score 135.5;  DB 23;  Length 530;
  Best Local Similarity   21.8%;  Pred. No. 0.008;
  Matches   79;  Conservative   59;  Mismatches  110;  Indels  115;  Gaps   18;

Qy          7 SEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQ----------L 55
              |  |   : || :   |:|:         |  | :||:  : :|  :
Db         93 SSQQATAKQSNASSDVEVEEK---------ETSVSKEDTDHEEKASNEDVTKAVDITTPK 143

Qy         56 GGRPG--PEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEV---E 110
                |  |   : |  :|::   : | |    :     ::|:           |   ||   :
Db        144 AARRGRKRKAEKQVETEEAGVVTTATASVNLKVSPKRGR----------PAATEVKIPK 192

Qy        111 PTPEAELLAQPCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKL 170
              |   ::: ||| ||  ::  :       ||:  ::: | ||     :| |
Db        193 PRGRPKMVKQPC-PSESDII------TEEDKSKKKGQEGK-----QPKK---------- 229

Qy        171 TWEEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTG---LYSK 227
              : || :| |                          :|  ::||| | |   : ||
Db        230 ---QPKKDEEGQK-------------------------EEDKPRKEPDKKEGKKEVESK 260

Qy        228 RA-AAKSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSETYER------YHTESLQ 280
              |   ||:  ||  |  ||| :::|     ||       |:|   : |      :  |:
Db        261 RKNLAKTGVTSTSDSEEEGDDQEGEKKRKGG-------RNFQTAHRRNMLKGQHEKEAAD 313

Qy        281 NMSKQELIKEYLELEKCLSRMEDENNR---LRLESKRLGGDDARVRELELEL-DRLRAEN 336
              |||   | :| |:   : :||  :    ::| ||    |:|:: :  |: : |: :|
Db        314 RKRKQE---EQMETEQ---QNKDEGKKPEVKKVEKKRETSMDSRLQRIHAEIKNSLKIDN 367
```

Qy 337 LQL 339
Db 368 LDV 370

Search completed: February 5, 2004, 13:28:25
Job time : 83 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: February 5, 2004, 13:27:09 ; Search time 28 Seconds
(without alignments)
542.486 Million cell updates/sec

Title: US-09-972-758A-2
Perfect score: 1910
Sequence: 1 MAEPFLSEYQHQPQTSNCTG.....LTENELHRQQERAPLSKFGD 359

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	134.5	7.0	939	4	US-09-252-991A-19109	Sequence 19109, A
2	133	7.0	538	4	US-09-309-572-12	Sequence 12, Appl
3	133	7.0	1312	4	US-09-554-572-26	Sequence 26, Appl
4	133	7.0	1737	4	US-09-309-572-13	Sequence 13, Appl
5	131	6.9	1898	1	US-08-056-200-94	Sequence 94, Appl
6	131	6.9	1898	2	US-08-800-644-94	Sequence 94, Appl
7	130	6.8	611	4	US-09-216-393B-81	Sequence 81, Appl
8	125	6.5	754	3	US-09-214-564A-2	Sequence 2, Appli
9	123.5	6.5	557	4	US-08-979-608A-5	Sequence 5, Appli
10	123	6.4	1829	3	US-09-157-420-1	Sequence 1, Appli
11	122	6.4	723	1	US-07-814-964-11	Sequence 11, Appl
12	122	6.4	723	1	US-08-258-442-11	Sequence 11, Appl
13	122	6.4	723	1	US-08-328-809-6	Sequence 6, Appli
14	122	6.4	723	4	US-08-866-840-6	Sequence 6, Appli
15	122	6.4	723	5	PCT-US92-11107-11	Sequence 11, Appl
16	121	6.3	538	3	US-09-370-368-9	Sequence 9, Appli
17	120	6.3	1780	1	US-08-769-309A-5	Sequence 5, Appli
18	120	6.3	1780	3	US-08-994-570-5	Sequence 5, Appli
19	118	6.2	820	4	US-09-252-991A-23346	Sequence 23346, A
20	118	6.2	830	4	US-09-562-737-39	Sequence 39, Appl
21	116.5	6.1	369	4	US-09-252-991A-20245	Sequence 20245, A
22	116	6.1	530	4	US-08-979-608A-8	Sequence 8, Appli
23	116	6.1	2972	4	US-09-579-181-2	Sequence 2, Appli
24	116	6.1	3118	4	US-09-579-181-1	Sequence 1, Appli
25	115.5	6.0	1596	4	US-08-978-277A-4	Sequence 4, Appli
26	114.5	6.0	245	4	US-09-252-991A-29400	Sequence 29400, A
27	114.5	6.0	1958	1	US-07-945-283-2	Sequence 2, Appli
28	114	6.0	683	6	5210183-3	Patent No. 5210183
29	113	5.9	380	4	US-09-252-991A-22319	Sequence 22319, A
30	113	5.9	8991	4	US-08-714-741-32	Sequence 32, Appl
31	112.5	5.9	599	3	US-08-556-419-22	Sequence 22, Appl
32	112.5	5.9	629	3	US-08-556-419-23	Sequence 23, Appl
33	112.5	5.9	764	4	US-09-370-838-67	Sequence 67, Appl
34	112	5.9	802	3	US-09-156-316-1	Sequence 1, Appli
35	111	5.8	1162	2	US-08-728-323A-2	Sequence 2, Appli
36	111	5.8	1162	4	US-09-298-568-2	Sequence 2, Appli
37	111	5.8	1162	4	US-09-410-399-2	Sequence 2, Appli
38	110	5.8	576	2	US-08-533-306A-2	Sequence 2, Appli
39	110	5.8	576	2	US-08-742-923A-2	Sequence 2, Appli
40	108.5	5.7	634	4	US-09-252-991A-24400	Sequence 24400, A
41	108.5	5.7	657	3	US-08-893-852A-3	Sequence 3, Appli
42	108.5	5.7	657	3	US-08-821-818-3	Sequence 3, Appli
43	108.5	5.7	657	4	US-09-052-753B-3	Sequence 3, Appli
44	108	5.7	631	4	US-09-252-991A-20063	Sequence 20063, A
45	108	5.7	671	4	US-09-439-313-380	Sequence 380, App

ALIGNMENTS

RESULT 1
US-09-252-991A-19109
; Sequence 19109, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19109
; LENGTH: 939
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19109

Query Match 7.0%; Score 134.5; DB 4; Length 939;
Best Local Similarity 20.8%; Pred. No. 0.0037;
Matches 88; Conservative 44; Mismatches 146; Indels 145; Gaps 17;

```
Qy          12 QPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGG--------RPGPEG 63
               ||:    |         | | ||||:  | |    |   ||   :||       :| |
Db          47 QPEADRRTA-----RRLRPRRPPGSRRRQPAFRLRQSGRALGPVGGVPLLVPAAQPQPRA 101

Qy          64 EGSLESQP-PPLQTQACPESSCLREG-------------EKGQNGDDSSAGGDFP-PPA 107
                  : :|  ||        ||  |                | : | | :    | |||
Db         102 YRPAQQRPAQPLAADPAGRSSVARSAGTRRDLRALWRRRRRAGGSDDASRSRRSHPHPPA 161

Qy         108 EVEPTPEAEL--------------------LAQPCHDSEA-----SKLGAPA------- 134
                : |  :  |                   |:|  |  |     |:| | |
Db         162 RLYPRADRRLPPSGADGRHEHPRRRSSGKLAPARPRPDGAAGRGDVSQLAAAAQSRSHPA 221

Qy         135 ----------------------AGGEEEWGQQQRQ-LGKKKHRRRPSKKK---RHWKPYY 168
                                    ||:  ||  || ||: :  |  :  :   : ||  |
Db         222 QFRSRSRAGRRARPADLRPSAGIGGDSPAGQPARQGLGRARGCRESAMSRDDVQRWKDKY 281

Qy         169 KLTWEEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKR 228
                  |::::   :   |   :|                                ||
Db         282 LENIEQQERLQRRWDARIDLLR-----------------------------RGLVRSS 310

Qy         229 AAAKSDDTSDDDFMEE------GGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNM 282
                ||:  | : |  |:|       : | |  |:       |::   :: :|    : ||:
Db         311 LAAEGSDKAVDQCMKELREILRRDDMDAGLSGL----IPRLEKAVLDSEQRRQQRTQQNI 366
```

```
Qy        283 -SKQELIKEYLELEKCLSRMEDENNRLRLESKRLGGD-DARVRE------LELELDRLRA 334
              :  || :: | |        |    ||   |:   | : | |:      |  || ||:
Db        367 DALGELAQQLLAL--------DLPRELRKPLKQFARDIEERARQSREIPILLSELSRLQR 418

Qy        335 ENL 337
              : |
Db        419 QAL 421


RESULT 2
US-09-309-572-12
; Sequence 12, Application US/09309572
; Patent No. 6440730
; GENERAL INFORMATION:
;  APPLICANT: Heinrich-Pette-Institut
;  TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
;  FILE REFERENCE: P50489
;  CURRENT APPLICATION NUMBER: US/09/309,572
;  CURRENT FILING DATE: 1999-05-11
;  EARLIER APPLICATION NUMBER: DE 198 56 463
;  EARLIER FILING DATE: 1998-11-26
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 538
;   TYPE: PRT
;   ORGANISM: Moloney murine leukemia virus
;   FEATURE:
;   OTHER INFORMATION: gag protein
US-09-309-572-12

  Query Match             7.0%;  Score 133;  DB 4;  Length 538;
  Best Local Similarity   20.9%;  Pred. No. 0.0024;
  Matches   89;  Conservative   40;  Mismatches  142;  Indels  154;  Gaps   19;

Qy         30 PERPPGA-------EERVPEEDSRWQSRAFPQLGGRPGPE---GEGS-----LESQPPPL 74
              |  || |           |   | : :   | || :| |:      |       |  |||
Db        107 PPLPPSAPSLPLEPPRSTPPRSSLYPALT-PSLGAKPKPQVLSDSGGPLIDLLTEDPPPY 165

Qy         75 QTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELL----AQPCHDSEASKL 130
              :    | |   |:|    || :::  |: |     :|:| |  |       |  ||  |:
Db        166 RDPRPPPSD--RDG----NGGEATPAGEAP-----DPSPMASRLRGRREPPVADSTTSQA 214

Qy        131 GAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDEKQSLRASRIR 190
              |||           |: ::    |    :||          | |      : |
Db        215 FPLRAGGN----------GQLQYWPFSSSDLYNWK-------NNNPSFSEDPGKLTALIE 257

Qy        191 AEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTG------LYSKRAAAKSDD--------- 235
              : :          | |  ||  |   | ||      |   | | : ||
Db        258 SVLI----------THQPTWDDCQQLLGTLLTGEEKQRVLLEARKAVRGDDGRPTQLPNE 307

Qy        236 --------TSDDDFMEEGGEE-------------------------DGGSDGMGGDGSE 261
                     | |:  : |                           | : |     |
Db        308 VDAAFPLERPDWDYTTQAGRNHLVHYRQLLLAGLQNAGRSPTNLAKVKGITQGPNESPSA 367

Qy        262 FLQRDFSETYERY-----------------------------HTESLQNMSKQELIKE 290
              ||:|   | | ||                           | |:| :  :|::
Db        368 FLER-LKEAYRRYTPYDPEDPGQETNVSMSFIWQSAPDIGRKLERLEDLKNKTLGDLVR- 425

Qy        291 YLELEKCLSRMEDENNRLRLESKRLGGDDARV-RELELELDRLRAENLQLLTENELHRQQ 349
                | ||  :: |   |          : |: || | :| | |: |   | :  | :
Db        426 --EAEKIFNKRETPEER----------EERIRRETEEKEERRRTEDEQKEKERDRRRHR 472

Qy        350 ERAPL 354
              | : |
Db        473 EMSKL 477


RESULT 3
US-09-554-572-26
; Sequence 26, Application US/09554572
; Patent No. 6573091
; GENERAL INFORMATION:
;  APPLICANT: NATURE TECHNOLOGY, INC.
;  TITLE OF INVENTION: CHIMERIC VIRAL PACKAGING SIGNAL WITHOUT GAG GENE
;  TITLE OF INVENTION:  SEQUENCES
;  FILE REFERENCE: 228.00030201
;  CURRENT APPLICATION NUMBER: US/09/554,572
;  CURRENT FILING DATE:  2000-09-18
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
;   LENGTH: 1312
;   TYPE: PRT
;   ORGANISM: Murine leukemia virus
US-09-554-572-26

  Query Match             7.0%;  Score 133;  DB 4;  Length 1312;
  Best Local Similarity   20.9%;  Pred. No. 0.0078;
  Matches   89;  Conservative   40;  Mismatches  142;  Indels  154;  Gaps   19;

Qy         30 PERPPGA-------EERVPEEDSRWQSRAFPQLGGRPGPE---GEGS-----LESQPPPL 74
              |  || |           |   | : :   | || :| |:      |       |  |||
Db        217 PPLPPSAPSLPLEPPRSTPPRSSLYPALT-PSLGAKPKPQVLSDSGGPLIDLLTEDPPPY 275

Qy         75 QTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELL----AQPCHDSEASKL 130
              :    | |   |:|    || :::  |: |     :|:| |  |       |  ||  |:
Db        276 RDPRPPPSD--RDG----NGGEATPAGEAP-----DPSPMASRLRGRREPPVADSTTSQA 324

Qy        131 GAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDEKQSLRASRIR 190
              |||           |: ::    |    :||          | |      : |
Db        325 FPLRAGGN----------GQLQYWPFSSSDLYNWK-------NNNPSFSEDPGKLTALIE 367

Qy        191 AEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTG------LYSKRAAAKSDD--------- 235
              : :          | |  ||  |   | ||      |   | | : ||
Db        368 SVLI----------THQPTWDDCQQLLGTLLTGEEKQRVLLEARKAVRGDDGRPTQLPNE 417

Qy        236 --------TSDDDFMEEGGEE-------------------------DGGSDGMGGDGSE 261
                     | |:  : |                           | : |     |
Db        418 VDAAFPLERPDWDYTTQAGRNHLVHYRQLLLAGLQNAGRSPTNLAKVKGITQGPNESPSA 477

Qy        262 FLQRDFSETYERY-----------------------------HTESLQNMSKQELIKE 290
              ||:|   | | ||                           | |:| :  :|::
Db        478 FLER-LKEAYRRYTPYDPEDPGQETNVSMSFIWQSAPDIGRKLERLEDLKNKTLGDLVR- 535

Qy        291 YLELEKCLSRMEDENNRLRLESKRLGGDDARV-RELELELDRLRAENLQLLTENELHRQQ 349
                | ||  :: |   |          : |: || | :| | |: |   | :  | :
Db        536 --EAEKIFNKRETPEER----------EERIRRETEEKEERRRTEDEQKEKERDRRRHR 582

Qy        350 ERAPL 354
              | : |
Db        583 EMSKL 587


RESULT 4
US-09-309-572-13
; Sequence 13, Application US/09309572
; Patent No. 6440730
; GENERAL INFORMATION:
;  APPLICANT: Heinrich-Pette-Institut
;  TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
;  FILE REFERENCE: P50489
;  CURRENT APPLICATION NUMBER: US/09/309,572
;  CURRENT FILING DATE: 1999-05-11
;  EARLIER APPLICATION NUMBER: DE 198 56 463
;  EARLIER FILING DATE: 1998-11-26
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 1737
;   TYPE: PRT
;   ORGANISM: Moloney murine leukemia virus
;   FEATURE:
;   OTHER INFORMATION: gag-pol protein
US-09-309-572-13
```

```
  Query Match             7.0%;  Score 133;  DB 4;  Length 1737;
  Best Local Similarity   20.9%;  Pred. No. 0.011;
  Matches   89;  Conservative   40;  Mismatches  142;  Indels  154;  Gaps   19;

Qy         30 PERPPGA-------EERVPEEDSRWQSRAFPQLGGRPGPE---GEGS-----LESQPPPL 74
              |  || |           |   | : :   | || :| |:     |       |   |||
Db        107 PPLPPSAPSLPLEPPRSTPPRSSLYPALT-PSLGAKPKPQVLSDSGGPLIDLLTEDPPPY 165

Qy         75 QTQACPESSCLREGEKGQNGDDSSAGGDFPPPPAEVEPTPEAELL----AQPCHDSEASKL 130
              :   | |   |:|    || :::  |: |     :|:| |  |       |  || |:
Db        166 RDPRPPPSD--RDG----NGGEATPAGEAP-----DPSPMASRLRGRREPPVADSTTSQA 214

Qy        131 GAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDEKQSLRASRIR 190
                 |||           |: ::    |    :||           | |      : |
Db        215 FPLRAGGN----------GQLQYWPFSSSDLYNWK-------NNNPSFSEDPGKLTALIE 257

Qy        191 AEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTG------LYSKRAAAKSDD--------- 235
              : :           | |   ||  |    | ||      |   | | : ||
Db        258 SVLI----------THQPTWDDCQQLLGTLLTGEEKQRVLLEARKAVRGDDGRPTQLPNE 307

Qy        236 --------TSDDDFMEEGGEE-------------------------DGGSDGMGGDGSE 261
                      | |:  : |                           | : |      |
Db        308 VDAAFPLERPDWDYTTQAGRNHLVHYRQLLLAGLQNAGRSPTNLAKVKGITQGPNESPSA 367

Qy        262 FLQRDFSETYERY-------------------------------HTESLQNMSKQELIKE 290
              ||:|   | | ||                             | |:| :  :|::
Db        368 FLER-LKEAYRRYTPYDPEDPGQETNVSMSFIWQSAPDIGRKLERLEDLKNKTLGDLVR- 425

Qy        291 YLELEKCLSRMEDENNRLRLESKRLGGDDARV-RELELELDRLRAENLQLLTENELHRQQ 349
              | ||  :: |    |              : |: || | : :| | |: |    | :  | :
Db        426 --EAEKIFNKRETPEER----------EERIRRETEEKEERRRTEDEQKEKERDRRRHR 472

Qy        350 ERAPL 354
              | : |
Db        473 EMSKL 477
```

RESULT 5
US-08-056-200-94
; Sequence 94, Application US/08056200
; Patent No. 5616500
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; TITLE OF INVENTION: Mehods of Using Same
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/056,200
; FILING DATE: 30-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fedrick, Michael F.
; REGISTRATION NUMBER: 36,799
; REFERENCE/DOCKET NUMBER: NIH054.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502

; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1898 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-056-200-94

```
  Query Match             6.9%;  Score 131;  DB 1;  Length 1898;
  Best Local Similarity   22.4%;  Pred. No. 0.019;
  Matches   79;  Conservative   54;  Mismatches  146;  Indels   74;  Gaps   13;

Qy         31 ERPPGAEER-------VPEEDSRWQSRAFPQLGGRPGPEGEGSLESQPPPLQTQACPESS 83
              ||   ||||        :|||: : |        ||   | |  |:       | |     :
Db        813 ERQLRAEERQQREQRFLPEEEEKEQR-------GRQRRREREKELQFLEEEEQLQRRERAQ 865

Qy         84 CLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAE-------LLAQPCHDSEASKLGAPAAG 136
               |:| | |   |       :     :      | |        | |:|     :  |
Db        866 QLQEEEDGLQEDQERRRQEQRRDQKWRWQLEEERKRRRHTLYAKPALQEQLRKEQQLLQE 925

Qy        137 GEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDEKQSLRASRIRAEMFAK 196
              |||  :::|:       :||  :::| ::   :|  ||::   |::  |  : |   : |
Db        926 EEEELQREERE------KRRRQEQERQYREEEQLQQEEEQLLREEREKRRRQERERQYRK 979

Qy        197 GQPVAPYNTTQFLMDDHDQ---EEPDLKTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSD 253
               :          |    :|   |||: :     :|   :     :::  :|
Db        980 DKK---------LQQKEEQLLGEEPEKR-----RRQEREKKYREEEELQQE--------- 1016

Qy        254 GMGGDGSEFLQRDFSETYER------YHTESLQNMSKQELIKEYLELEKCLSRMED--EN 305
                    | | |:  |   |       |  :      :::|::|  |  :   |      |
Db       1017 ------EEQLLREEREKRRRQEWERQYRKKDELQQEEEQLLREEREKRRLQERERQYREE 1070

Qy        306 NRLRLESKRLGGDDARVR---ELELEL---DRLRAENLQLLTEN-ELHRQQER 351
                |: | ::| |::   |   ||| :    : |: |  |||  |  | |:|||
Db       1071 EELQQEEEQLLGEERETRRRQELERQYRKEEELQQEEEQLLREEPEKRRRQER 1123
```

RESULT 6
US-08-800-644-94
; Sequence 94, Application US/08800644
; Patent No. 5958752
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; TITLE OF INVENTION: Mehods of Using Same
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,644
; FILING DATE: 14-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/056,200
; FILING DATE: 30-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fedrick, Michael F.
; REGISTRATION NUMBER: 36,799
; REFERENCE/DOCKET NUMBER: NIH054.001A

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (714) 760-0404
;      TELEFAX:  (714) 760-9502
;  INFORMATION FOR SEQ ID NO:  94:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1898 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-800-644-94

  Query Match             6.9%;  Score 131;  DB 2;  Length 1898;
  Best Local Similarity   22.4%;  Pred. No. 0.019;
  Matches   79;  Conservative   54;  Mismatches  146;  Indels   74;  Gaps   13;

Qy         31 ERPPGAEER-------VPEEDSRWQSRAFPQLGGRPGPEGEGSLESQPPPLQTQACPESS 83
              ||   ||||       :|||: : |       ||   | |  |:        | |    :
Db        813 ERQLRAEERQQREQRFLPEEEEKEQR-------GRQRREREKELQFLEEEEQLQRRERAQ 865

Qy         84 CLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAE-------LLAQPCHDSEASKLGAPAAG 136
               |:| | |   |       :     :     | |       | |:|    :  |
Db        866 QLQEEEDGLQEDQERRRQEQRRDQKWRWQLEEERKRRRHTLYAKPALQEQLRKEQQLLQE 925

Qy        137 GEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDEKQSLRASRIRAEMFAK 196
              |||  :::|:      :||  :::| ::   :|  ||::    |::  |  : |   : |
Db        926 EEEELQREERE------KRRRQEQERQYREEEQLQQEEEQLLREEREKRRRQERERQYRK 979

Qy        197 GQPVAPYNTTQFLMDDHDQ---EEPDLKTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSD 253
               :          |   :|   |||: :     :|    :     :::  :|
Db        980 DKK---------LQQKEEQLLGEEPEKR-----RRQEREKKYREEEELQQE--------- 1016

Qy        254 GMGGDGSEFLQRDFSETYER------YHTESLQNMSKQELIKEYLELEKCLSRMED--EN 305
                  | | |:  |   |       |  :       :::|::|  |  :   |      |
Db       1017 ------EEQLLREEREKRRRQEWERQYRKKDELQQEEEQLLREEREKRRLQERERQYREE 1070

Qy        306 NRLRLESKRLGGDDARVR---ELELEL---DRLRAENLQLLTEN-ELHRQQER 351
                |: | ::| |::   |   ||| :    : |: |  ||| |  |  |:|||
Db       1071 EELQQEEEQLLGEERETRRRQELERQYRKEEELQQEEEQLLREEPEKRRRQER 1123


RESULT 7
US-09-216-393B-81
; Sequence 81, Application US/09216393B
; Patent No. 6514694
; GENERAL INFORMATION:
;  APPLICANT: Milhausen, Michael James
;  TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND USES THER
;  FILE REFERENCE: TX-1-C2
;  CURRENT APPLICATION NUMBER: US/09/216,393B
;  CURRENT FILING DATE:  1998-12-18
;  PRIOR APPLICATION NUMBER: 08/994,825
;  PRIOR FILING DATE: 1997-12-19
;  NUMBER OF SEQ ID NOS: 366
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 81
;   LENGTH: 611
;   TYPE: PRT
;   ORGANISM: Toxoplasma gondii
US-09-216-393B-81

  Query Match             6.8%;  Score 130;  DB 4;  Length 611;
  Best Local Similarity   22.2%;  Pred. No. 0.0052;
  Matches   82;  Conservative   58;  Mismatches  155;  Indels   74;  Gaps   15;

Qy         13 PQTSNCTGAAAVQEELN-PERPPGAEERVP-----EEDSRWQSRAFP----QLGGRPGPE 62
              |: |       : |  |: ||:  |:     |      :  : : | | |     :|   ||
Db        189 PRFSKSDVCCSPQARLSLPEQSLGSSPSSPISVTNDVYALFDSSASPLHAGELSSLPGAV 248

Qy         63 GEGSLESQPP----PLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELL 118
                  |    | : ||   ||                      | |: || :
Db        249 SASERLLTAPAEIGPSASSACLSVSC-------------------GPGEMSPTADTTR- 287

Qy        119 AQPCHDSEASKLGAPAAGGE----EEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEE 174
                ||:|  :         |    ||  :::|:: ::| | |  :::|      :   ||
Db        288 ----HDAEERERRRAEEEKERERQEEEERERRRVEEEKERERQEEEERE----RRRVEEE 339

Qy        175 K---KKFDEKQSLRA----SRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSK 227
              |   :: ||::  |      :| | |   : : |     |          ::|| : :   |
Db        340 KARQREEDERERRRVEEEKARQREEEERERRRVEEEKARQ----REEEEERERRRVEEEK 395

Qy        228 RAAAKSDDTSDDDFMEE----GGEEDGGSDGMGGDGSEFLQRDFSETYE--RYHTESLQN 281
                 : ::  :  :||       ||:   :|   :  :  ||:  |  |  |   | :
Db        396 ARQREEEEERERRRVEEEKARQREEEEEREGRRVEEEKARQREEEEEREGRRVEEEKARQ 455

Qy        282 MSKQELIKEYLELEKCLSRMEDENNRLRLESKRLGGDDARVRELELELDRLRAENLQLLT 341
               ::|  :  :| ||    | |:|  | |:| ::       | |:  | | :| | |
Db        456 REEEERERRRVEEEKERERQEEERERRRVEEEK-----ERERQEEEERERRRVEE----- 505

Qy        342 ENELHRQQE 350
              | |  ||:|
Db        506 EKERERQEE 514


RESULT 8
US-09-214-564A-2
; Sequence 2, Application US/09214564A
; Patent No. 6150515
; GENERAL INFORMATION:
;  APPLICANT:  Sharp, Phillip A.
;  APPLICANT:  Zhou, Qiang
;  TITLE OF INVENTION:  TAT-SF: Cofactor For Stimulation Of Transcriptional
;  TITLE OF INVENTION:  Elongation By HIV-1 TAT
;  FILE REFERENCE:  M0656/7042
;  CURRENT APPLICATION NUMBER: US/09/214,564A
;  CURRENT FILING DATE:  1999-08-18
;  PRIOR APPLICATION NUMBER:  US 60/021,218
;  PRIOR FILING DATE:  1996-07-03
;  PRIOR APPLICATION NUMBER:  US 60/033,152
;  PRIOR FILING DATE:  1996-12-13
;  PRIOR APPLICATION NUMBER:  PCT/US97/11713
;  PRIOR FILING DATE:  1997-07-03
;  NUMBER OF SEQ ID NOS:  5
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 754
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-214-564A-2

  Query Match             6.5%;  Score 125;  DB 3;  Length 754;
  Best Local Similarity   21.5%;  Pred. No. 0.018;
  Matches   82;  Conservative   50;  Mismatches  166;  Indels   84;  Gaps   17;

Qy          9 YQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGG--RPGPEGEGS 66
              :  | ||                |   |     | :: :::     : ||     |     :
Db        401 FSEHPSTSKMNA---------QETATGMAFEEPIDEKKPEK---TEDGGEFEEGASENNA 448

Qy         67 LESQPPPLQTQACP----ESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAE----LL 118
               || |     : ||    |  | : | :|       |  |: |     | :|: |     |
Db        449 KESSPEKEAEEGCPEKESEEGCPKRGFEGSCSQKESEEGN-PVRGSEEDSPKKESKKKTL 507

Qy        119 AQPCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKF 178
                | ::  :|         | |      ::|  |       :|           |  :
Db        508 KNDCEENGLAKESEDDLNKESE-----EEVGPTKESEEDDSEK-----------ESDEDC 551

Qy        179 DEKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQE--EPDLKTGLYSKRAAAKSDDT 236
               ||||    |     |                  |  | |:|  | :|   :  |    : :|:
Db        552 SEKQSEDGSEREFE------------ENGLEKDLDEEGSEKELHENVLDKE--LEENDS 596

Qy        237 SDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSE----------TYER-YHTESLQNMSKQ 285
               : :|     |:||     :  :|||   |:| |           |||: :  || :   ::
Db        597 ENSEF-----EDDGSEKVLDEEGSE---REFDEDSDEKEEEEDTYEKVFDDESDEKEDEE 648

Qy        286 ELIKEYLELEKCLSRMEDENNRLRLESKRLGGDDARVRELELELDRLRAE---NLQLLTE 342
                ::  ||      :   | : :|  ||      :||  :|:|   ::|  :   | :|  |
Db        649 YADEKGLEAADKKAEEGDADEKLFEESDDKEDEDADGKEVEDADEKLFEDDDSNEKLFDE 708
```

```
Qy        343 NE------LHRQQERAPLSKFG 358
              |      ||  |  ||
Db        709 EEDSSEKLFDDSDERGTLGGFG 730


RESULT 9
US-08-979-608A-5
; Sequence 5, Application US/08979608A
; Patent No. 6355451
;   GENERAL INFORMATION:
;        APPLICANT: Lees, Ann M.
;                   Lees, Robert S.
;                   Law, Simon W.
;                   Arjona, Anibal A.
;        TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
;                            BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
;                            TREATING ATHEROSCLEROSIS
;        NUMBER OF SEQUENCES: 42
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Fish & Richardson P.C.
;             STREET: 225 Franklin Street
;             CITY: Boston
;             STATE: MA
;             COUNTRY: USA
;             ZIP: 02110-2804
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ for Windows Version 2.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/979,608A
;             FILING DATE: 26-No. 6355451-1997
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US 60/048,547
;             FILING DATE: 03-JUN-1997
;             APPLICATION NUMBER: US 60/031,930
;             FILING DATE: 27-NOV-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Myers, Louis
;             REGISTRATION NUMBER: 35,965
;             REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 617/542-5070
;             TELEFAX: 617/542-8906
;   INFORMATION FOR SEQ ID NO: 5:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 557 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        FRAGMENT TYPE: internal
;        SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-08-979-608A-5

  Query Match             6.5%;  Score 123.5;  DB 4;  Length 557;
  Best Local Similarity  22.6%;  Pred. No. 0.017;
  Matches   91;  Conservative   53;  Mismatches  123;  Indels  135;  Gaps   20;

Qy         20 GAAAVQEELNPERPPGAEERVPE---------EDSRWQSRAFPQLGGRPGPEGEGSLESQ 70
              |||   :: ||:  ||  |   |           :|    |  |  ||  |||  :  :|
Db          9 GAA---KQPNPKSSPGQPEAGAEGAQGRPGRPAPAREAEGASSQAPGR--PEGAQAKTAQ 63

Qy         71 PPPL---------QTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELLAQP 121
              | |          | :    : |: :  :|  |:|   |           || || |
Db         64 PGALCDVSEELSRQLEDILSTYCV-DNNQGAPGEDGVQG---------EP-PEPE----- 107

Qy        122 CHDSEASKL----------GAPAAGGEEEWGQQQ---------RQLGKKKHRRRPSKKKR 162
                |:| |:          | |   ||:|  : :          ::| : |||   |||
Db        108 --DAEKSRAYVARNGEPEPGTPVVNGEKETSKAEPGTEEIRTSDEVGDRDHRRPQEKKKA 165

Qy        163 HW---------------KPYYKLTWEEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNTT 206
                                |  ||    ||  :  :  | |: : ::  | |       :
Db        166 KGLGKEITLLMQTLNTLSTPEEKLAALCKKYAELLEEHRNSQKQMKLLQKKQ-------S 218

Qy        207 QFLMDDHDQEEPDLKTGLYSKRAAAKSDDTS--------DDDFMEEGGEEDGGSDGMGGD 258
              | :      :| |   | :||   |:|   |         :   |||
Db        219 QLV------QEKDHLRGEHSKAILARSKLESLCRELQRHNRSLKEEG------------- 259

Qy        259 GSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCLSRMEDENNRLRLESKRLGGD 318
                :||   |  :|    |   |:  ::    |::|    :  : |::|| |: |
Db        260 ----VQRAREEEEKRKEVTSHFQMTLNDI---QLQME----QHNERNSKLRQENMELA-- 306

Qy        319 DARVRELELELDRLRAENL----------QLLTENELHRQQE 350
                |:::| :|   || |::          | | : :| : ||
Db        307 -ERLKKL-IEQYELREEHIDKVFKHKDLQQQLVDAKLQQAQE 346


RESULT 10
US-09-157-420-1
; Sequence 1, Application US/09157420
; Patent No. 6180760
; GENERAL INFORMATION:
;  APPLICANT: TAKAI, Yoshimi
;  APPLICANT: NAKANISHI, Hiroyuki
;  APPLICANT: MANDAI, Kenji
;  APPLICANT: WADA, Manabu
;  APPLICANT: OBAISHI, Hiroshi
;  TITLE OF INVENTION: ACTIN FILAMENT-BINDING PROTEIN "L-AFADIN"
;  FILE REFERENCE: 98-1042*/LC(WMC)/653
;  CURRENT APPLICATION NUMBER: US/09/157,420
;  CURRENT FILING DATE: 1998-09-21
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 1829
;   TYPE: PRT
;   ORGANISM: rat
US-09-157-420-1

  Query Match             6.4%;  Score 123;  DB 3;  Length 1829;
  Best Local Similarity  19.9%;  Pred. No. 0.087;
  Matches   83;  Conservative   59;  Mismatches  157;  Indels  118;  Gaps   15;

Qy          9 YQHQPQTSNCTGAAAV--------QEELNPERPPGAEERVPE-------EDSRWQSRAFP 53
              |: :|      :  |::        ||||  |:    |    |       :   | :::
Db       1259 YEEKPHMHTESDHASIAIQRVTRSQEELREEKVYQLERHRVESGMDRKCDSDMWINQS-- 1316

Qy         54 QLGGRPGPEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTP 113
                  |   | :           | |:  : |             | :  || | |||
Db       1317 -----SSVESSTSSQEHLNHSSKSVTPASTLTKSGP----------GRWKTPAAVLPTP 1360

Qy        114 EAELLAQPCH----------------DSEASKLGAPAAGGEEEWGQQQRQLGKKKHRR 155
                |  ::||               | :  |  |     :   |  |:   : ::
Db       1361 VA--VSQPIRTDLPPPPPPPPAHYTSDFDGISMDLPLPPPPANQA-APQSAQVAAAERKK 1417

Qy        156 RPSKKKRHWKPYYKLTWE-EKKKFDEKQSLRASRIRAEMFAKGQPVA------------ 201
              |   :: : |   :|  | |:|: :::: |   | ::   |   |:|
Db       1418 REEHQRWYEKEKARLEEERERKRREQERKLGQMRTQSLNPASFSPLATQAKPEKPSTLQR 1477

Qy        202 PYNTTQFLMDDHDQEEP------DLKTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSDGM 255
              | |   : :   |::|      ||:    ||   :  |  | | :
Db       1478 PQETV--IRELQPQQQPRTIERRDLQYITISKEELSSGDSLSPDPW-------------- 1521

Qy        256 GGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCLSRMEDENNRLR------ 309
                  :||  |  |:     : :|    | ||  ||:    |  :|::|||
Db       1522 --------KRDAREKLEKQQQMHIVDM----LSKEIHELQNKGDRTAEESDRLRKLMLEW 1569

Qy        310 ------LESKRLGGDDARVRELELE----LDRLRAENLQLLTENELHRQQERAPLSK 356
                    |||:   ||    : :::    : || ||    | : |  |||:   : |
Db       1570 QFQKRLQESKQKDEDDDEEEDDDVDTMLIMQRLEAERRARLQDEERRRQQQLEEMRK 1626


RESULT 11
US-07-814-964-11
```

```
; Sequence 11, Application US/07814964
; Patent No. 5359047
;  GENERAL INFORMATION:
;    APPLICANT:  Donahue, Brian A.
;    APPLICANT:  Toney, Jeffrey H.
;    APPLICANT:  Bruhn, Suzanne L.
;    APPLICANT:  Pil, Pieter M.
;    APPLICANT:  Brown, Steven
;    APPLICANT:  Kellett, Patti
;    APPLICANT:  Essigmann, John M.
;    APPLICANT:  Lippard, Stephen J.
;    TITLE OF INVENTION:  DNA Structure Specific Recognition
;    TITLE OF INVENTION:  Protein and Uses Therefor
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  2 Militia Drive
;      CITY:  Lexington
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/814,964
;      FILING DATE:  19911226
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/539,906
;      FILING DATE:  18-JUN-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  MIT-4787AAA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  11
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  723 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Drosophila melanogaster
;    IMMEDIATE SOURCE:
;      CLONE:  Drosophila SSRP (predicted)
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  458..507
;      OTHER INFORMATION:  /label= Acidic
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  518..547
;      OTHER INFORMATION:  /label= Basic I
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  547..620
;      OTHER INFORMATION:  /label= HMG-box
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  632..649
;      OTHER INFORMATION:  /label= Basic II
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  657..723
;      OTHER INFORMATION:  /label= Mixed Charge
US-07-814-964-11
```

```
  Query Match             6.4%;  Score 122;  DB 1;  Length 723;
  Best Local Similarity   20.4%;  Pred. No. 0.031;
  Matches   61;  Conservative   55;  Mismatches  125;  Indels   58;  Gaps    9;

Qy         63 GEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPP---PAEVEPTPEAELLA 119
              |:   |::|     :   |:    | :   : |: |   || |    ::|    :: : :
Db        441 GDSDNENEPDAYLARLKAEAREKEEDDDDGDSDEESTDEDFKPNENESDVAEEYDSNVES 500

Qy        120 QPCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFD 179
                ||:||       ||: :  :::::   ||  ::  |:|:| |   :     ||| |
Db        501 DSDDDSDAS-----GGGGDSDGAKKKKE---KKSEKKEKKEKKH-KEKERTKKPSKKKKD 551

Qy        180 EKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHD------QEEPDLKTGLYSKR----- 228
              :  ||                    || |::  :|      :| | :|    :|:
Db        552 SGKPKRA-----------------TTAFMLWLNDTRESIKRENPGIKVTEIAKKGGEMW 593

Qy        229 ----------AAAKSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTE 277
                       ||||      |:      | | ||   |  |   ::      :: :|
Db        594 KELKDKSKWEDAAAKDKQRYHDEMRNYKPEAGGDSDNEKGGKSSKKRKTEPSPSKKANTS 653

Qy        278 SLQNMSKQELIKEYLELEKCLSRMEDENNR-LRLESKRLGGDDARVRELELELDRLRAE 335
                  |    |||:  :   |  ::::|   : :||     ||: :: : | :   :|
Db        654 GSGFKS-----KEYISDDDSTSSDDEKDNEPAKKKSKPPSDGDAKKKKAKSESEPEESE 707
```

```
RESULT 12
US-08-258-442-11
; Sequence 11, Application US/08258442
; Patent No. 5670621
;  GENERAL INFORMATION:
;    APPLICANT:  Donahue, Brian A.
;    APPLICANT:  Toney, Jeffrey H.
;    APPLICANT:  Bruhn, Suzanne L.
;    APPLICANT:  Pil, Pieter M.
;    APPLICANT:  Brown, Steven
;    APPLICANT:  Kellett, Patti
;    APPLICANT:  Essigmann, John M.
;    APPLICANT:  Lippard, Stephen J.
;    TITLE OF INVENTION:  DNA Structure Specific Recognition
;    TITLE OF INVENTION:  Protein and Uses Therefor
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  2 Militia Drive
;      CITY:  Lexington
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/258,442
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/539,906
;      FILING DATE:  18-JUN-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  MIT-4787AAA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  11:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  723 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
```

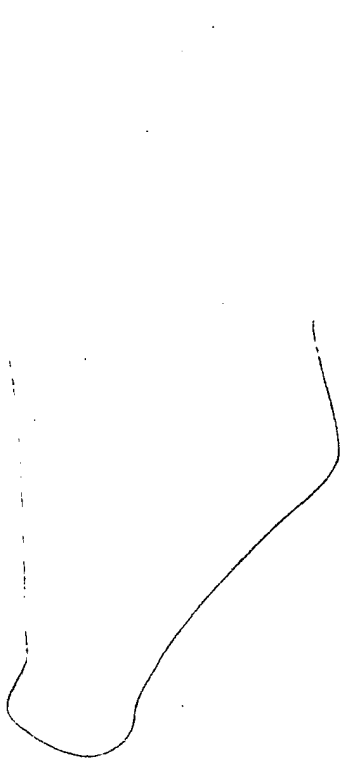
```
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Drosophila melanogaster
;    IMMEDIATE SOURCE:
;      CLONE:  Drosophila SSRP (predicted)
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  458..507
;      OTHER INFORMATION:  /label= Acidic
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  518..547
;      OTHER INFORMATION:  /label= Basic I
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  547..620
;      OTHER INFORMATION:  /label= HMG-box
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  632..649
;      OTHER INFORMATION:  /label= Basic II
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  657..723
;      OTHER INFORMATION:  /label= Mixed Charge
US-08-258-442-11

  Query Match             6.4%;  Score 122;  DB 1;  Length 723;
  Best Local Similarity   20.4%;  Pred. No. 0.031;
  Matches   61;  Conservative   55;  Mismatches  125;  Indels   58;  Gaps    9;

Qy         63 GEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPP---PAEVEPTPEAELLA 119
              |:   |::|     :   |:    | :   : |: |   || |    ::|    :: : :
Db        441 GDSDNENEPDAYLARLKAEAREKEEDDDDGDSDEESTDEDFKPNENESDVAEEYDSNVES 500

Qy        120 QPCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFD 179
                ||:||       ||: :  :::::   ||  ::  |:|:| |   :     ||| |
Db        501 DSDDDSDAS-----GGGGDSDGAKKKKE---KKSEKKEKKEKKH-KEKERTKKPSKKKKD 551

Qy        180 EKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHD------QEEPDLKTGLYSKR----- 228
               :  ||                  || |::  :|      :| | :|    :|:
Db        552 SGKPKRA-----------------TTAFMLWLNDTRESIKRENPGIKVTEIAKKGGEMW 593

Qy        229 -----------AAAKSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTE 277
                        ||||      |:       | | ||   | |   ::       :: :|
Db        594 KELKDKSKWEDAAAKDKQRYHDEMRNYKPEAGGDSDNEKGGKSSKKRKTEPSPSKKANTS 653

Qy        278 SLQNMSKQELIKEYLELEKCLSRMEDENNR-LRLESKRLGGDDARVRELELELDRLRAE 335
                  |     |||:  :   |  ::::|   : :||     ||: :: : | :   :|
Db        654 GSGFKS-----KEYISDDDSTSSDDEKDNEPAKKKSKPPSDGDAKKKKAKSESEPEESE 707


RESULT 13
US-08-328-809-6
; Sequence 6, Application US/08328809
; Patent No. 5705334
;  GENERAL INFORMATION:
;    APPLICANT:  Lippard, Stephen J.
;    APPLICANT:  Essigmann, John M.
;    APPLICANT:  Donahue, Brian A.
;    APPLICANT:  Toney, Jeffrey H.
;    APPLICANT:  Bruhn, Suzanne L.
;    APPLICANT:  Pil, Pieter M.
;    APPLICANT:  Brown, Steven
;    APPLICANT:  Kellett, Patti
;    TITLE OF INVENTION:  Uses For DNA Structure-Specific
;    TITLE OF INVENTION:  Recognition Proteins
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patent Administrator, Testa, Hurwitz & Thibeault
;      STREET:  53 State Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/328,809
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fenton, Gillian M.
;      REGISTRATION NUMBER:  36,508
;      REFERENCE/DOCKET NUMBER:  MIT-023 (5473/24)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-248-7000
;      TELEFAX:  617-248-7100
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  723 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Drosophila melanogaster
;    IMMEDIATE SOURCE:
;      CLONE:  Drosophila SSRP (predicted)
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  458..507
;      OTHER INFORMATION:  /label= Acidic
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  518..547
;      OTHER INFORMATION:  /label= Basic I
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  547..620
;      OTHER INFORMATION:  /label= HMG
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  632..649
;      OTHER INFORMATION:  /label= Basic II
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  657..723
;      OTHER INFORMATION:  /label= Mixed Charge
US-08-328-809-6

  Query Match             6.4%;  Score 122;  DB 1;  Length 723;
  Best Local Similarity   20.4%;  Pred. No. 0.031;
  Matches   61;  Conservative   55;  Mismatches  125;  Indels   58;  Gaps    9;

Qy         63 GEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPP---PAEVEPTPEAELLA 119
              |:   |::|     :   |:    | :   : |: |   || |    ::|    :: : :
Db        441 GDSDNENEPDAYLARLKAEAREKEEDDDDGDSDEESTDEDFKPNENESDVAEEYDSNVES 500

Qy        120 QPCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFD 179
                ||:||       ||: :  :::::   ||  ::  |:|:| |   :     ||| |
Db        501 DSDDDSDAS-----GGGGDSDGAKKKKE---KKSEKKEKKEKKH-KEKERTKKPSKKKKD 551

Qy        180 EKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHD------QEEPDLKTGLYSKR----- 228
               :  ||                  || |::  :|      :| | :|    :|:
Db        552 SGKPKRA-----------------TTAFMLWLNDTRESIKRENPGIKVTEIAKKGGEMW 593

Qy        229 -----------AAAKSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTE 277
                        ||||      |:       | | ||   | |   ::       :: :|
Db        594 KELKDKSKWEDAAAKDKQRYHDEMRNYKPEAGGDSDNEKGGKSSKKRKTEPSPSKKANTS 653

Qy        278 SLQNMSKQELIKEYLELEKCLSRMEDENNR-LRLESKRLGGDDARVRELELELDRLRAE 335
                  |     |||:  :   |  ::::|   : :||     ||: :: : | :   :|
Db        654 GSGFKS-----KEYISDDDSTSSDDEKDNEPAKKKSKPPSDGDAKKKKAKSESEPEESE 707
```

```
RESULT 14
US-08-866-840-6
; Sequence 6, Application US/08866840
; Patent No. 6475791
;  GENERAL INFORMATION:
;    APPLICANT:  Lippard, Stephen J.
;    APPLICANT:  Essigmann, John M.
;    APPLICANT:  Donahue, Brian A.
;    APPLICANT:  Toney, Jeffrey H.
;    APPLICANT:  Bruhn, Suzanne L.
;    APPLICANT:  Pil, Pieter M.
;    APPLICANT:  Brown, Steven
;    APPLICANT:  Kellett, Patti
;    TITLE OF INVENTION:  Uses For DNA Structure-Specific
;    TITLE OF INVENTION:  Recognition Proteins
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patent Administrator, Testa, Hurwitz & Thibeault
;      STREET:  53 State Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/866,840
;      FILING DATE:  02-JUN-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fenton, Gillian M.
;      REGISTRATION NUMBER:  36,508
;      REFERENCE/DOCKET NUMBER:  MIT-023 (5473/24)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-248-7000
;      TELEFAX:  617-248-7100
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  723 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Drosophila melanogaster
;    IMMEDIATE SOURCE:
;      CLONE:  Drosophila SSRP (predicted)
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  458..507
;      OTHER INFORMATION:  /label= Acidic
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  518..547
;      OTHER INFORMATION:  /label= Basic I
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  547..620
;      OTHER INFORMATION:  /label= HMG
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  632..649
;      OTHER INFORMATION:  /label= Basic II
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  657..723
;      OTHER INFORMATION:  /label= Mixed Charge
US-08-866-840-6
```

```
  Query Match             6.4%;  Score 122;  DB 4;  Length 723;
  Best Local Similarity   20.4%;  Pred. No. 0.031;
  Matches   61;  Conservative   55;  Mismatches  125;  Indels   58;  Gaps    9;

Qy         63 GEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPP---PAEVEPTPEAELLA 119
              |:   |::|      :   |:    | :   : |: |   || |    ::|    :: : :
Db        441 GDSDNENEPDAYLARLKAEAREKEEDDDDGDSDEESTDEDFKPNENESDVAEEYDSNVES 500

Qy        120 QPCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFD 179
                ||:||        ||: :  :::::   ||  ::  |:|:| |   :     ||| |
Db        501 DSDDDSDAS-----GGGGDSDGAKKKKE---KKSEKKEKKEKKH-KEKERTKKPSKKKKD 551

Qy        180 EKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHD------QEEPDLKTGLYSKR----- 228
              :  ||                  || |::  :|      :| | :|    :|:
Db        552 SGKPKRA-----------------TTAFMLWLNDTRESIKRENPGIKVTEIAKKGGEMW 593

Qy        229 ----------AAAKSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTE 277
                       ||||       |:       |  | ||   |  |   ::       :: :|
Db        594 KELKDKSKWEDAAAKDKQRYHDEMRNYKPEAGGDSDNEKGGKSSKKRKTEPSPSKKANTS 653

Qy        278 SLQNMSKQELIKEYLELEKCLSRMEDENNR-LRLESKRLGGDDARVRELELELDRLRAE 335
                  |     |||:  :   |  ::::|   : :||      ||: :: : | :   :|
Db        654 GSGFKS-----KEYISDDDSTSSDDEKDNEPAKKKSKPPSDGDAKKKKAKSESEPEESE 707
```

```
RESULT 15
PCT-US92-11107-11
; Sequence 11, Application PC/TUS9211107
;  GENERAL INFORMATION:
;    APPLICANT:  Donahue, Brian A.
;    APPLICANT:  Toney, Jeffrey H.
;    APPLICANT:  Bruhn, Suzanne L.
;    APPLICANT:  Pil, Pieter M.
;    APPLICANT:  Brown, Steven
;    APPLICANT:  Kellett, Patti
;    APPLICANT:  Essigmann, John M.
;    APPLICANT:  Lippard, Stephen J.
;    TITLE OF INVENTION:  DNA Structure Specific Recognition
;    TITLE OF INVENTION:  Protein and Uses Therefor
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  2 Militia Drive
;      CITY:  Lexington
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/11107
;      FILING DATE:  19921218
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/539,906
;      FILING DATE:  18-JUN-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  MIT-4787AAA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  723 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
```

```
;       ORGANISM:  Drosophila melanogaster
;     IMMEDIATE SOURCE:
;       CLONE:  Drosophila SSRP (predicted)
;     FEATURE:
;       NAME/KEY:  Domain
;       LOCATION:  458..507
;       OTHER INFORMATION:  /label= Acidic
;     FEATURE:
;       NAME/KEY:  Domain
;       LOCATION:  518..547
;       OTHER INFORMATION:  /label= Basic I
;     FEATURE:
;       NAME/KEY:  Domain
;       LOCATION:  547..620
;       OTHER INFORMATION:  /label= HMG-box
;     FEATURE:
;       NAME/KEY:  Domain
;       LOCATION:  632..649
;       OTHER INFORMATION:  /label= Basic II
;     FEATURE:
;       NAME/KEY:  Domain
;       LOCATION:  657..723
;       OTHER INFORMATION:  /label= Mixed Charge
PCT-US92-11107-11

  Query Match             6.4%;  Score 122;  DB 5;  Length 723;
  Best Local Similarity   20.4%;  Pred. No. 0.031;
  Matches   61;  Conservative   55;  Mismatches  125;  Indels   58;  Gaps    9;

Qy         63 GEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPP---PAEVEPTPEAELLA 119
              |:   |::|     :   |:    | :   : |: |·  || |    ::|    :: : :
Db        441 GDSDNENEPDAYLARLKAEAREKEEDDDDGDSDEESTDEDFKPNENESDVAEEYDSNVES 500

Qy        120 QPCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFD 179
                ||:||       ||: :  :::::   ||  ::  |:|:| |   :     ||| |
Db        501 DSDDDSDAS-----GGGGDSDGAKKKKE---KKSEKKEKKEKKH-KEKERTKKPSKKKKD 551

Qy        180 EKQSLRASRIRAEMPAKGQPVAPYNTTQFLMDDHD------QEEPDLKTGLYSKR----- 228
              :  ||                 || |::  :|      :| | :|    :|:
Db        552 SGKPKRA-----------------TTAFMLWLNDTRESIKRENPGIKVTEIAKKGGEMW 593

Qy        229 -----------AAAKSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTE 277
                        ||||     |:      |  | ||   |  |   ::       :: :|
Db        594 KELKDKSKWEDAAAKDKQRYHDEMRNYKPEAGGDSDNEKGGKSSKKRKTEPSPSKKANTS 653

Qy        278 SLQNMSKQELIKEYLELEKCLSRMEDENNR-LRLESKRLGGDDARVRELELELDRLRAE 335
                |     |||:  :   |  ::::|   : :||     ||: :: : | :   :|
Db        654 GSGFKS-----KEYISDDDSTSSDDEKDNEPAKKKSKPPSDGDAKKKKAKSESEPEESE 707


Search completed: February  5, 2004, 13:32:37
Job time : 29 secs
```



GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 5, 2004, 09:56:07 ; Search time 97 Seconds
(without alignments)
4914.369 Million cell updates/sec

Title: US-09-972-758A-1
Perfect score: 1080
Sequence: 1 atggccgagccattcttgtc.....tttccaagtttggagactag 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	69	6.4	7218	1	US-08-232-463-14	Sequence 14, Appl
	2	68.4	6.3	1926	4	US-09-249-585A-2	Sequence 2, Appli
	3	68.4	6.3	1926	4	US-09-410-399-3	Sequence 3, Appli
	4	68.4	6.3	2580	3	US-09-050-863-2	Sequence 2, Appli
	5	68.4	6.3	2580	4	US-09-359-081-2	Sequence 2, Appli
c	6	68.4	6.3	5452	2	US-09-130-114-1	Sequence 1, Appli
c	7	68.4	6.3	8705	4	US-09-647-344A-14	Sequence 14, Appl
	8	68.4	6.3	9600	3	US-08-910-647-1	Sequence 1, Appli
	9	68.4	6.3	9600	4	US-09-620-925-1	Sequence 1, Appli
	10	68.4	6.3	10596	1	US-07-884-811-15	Sequence 15, Appl
	11	68.4	6.3	10596	1	US-07-885-971-15	Sequence 15, Appl
	12	68.4	6.3	10596	1	US-08-087-783A-15	Sequence 15, Appl
	13	68.4	6.3	10596	1	US-08-194-088B-15	Sequence 15, Appl
	14	68.4	6.3	10596	2	US-08-194-087-15	Sequence 15, Appl
	15	68.4	6.3	10596	5	PCT-US93-04648-15	Sequence 15, Appl
c	16	67.4	6.2	1926	4	US-09-249-585A-4	Sequence 4, Appli
c	17	67.4	6.2	1931	2	US-09-130-114-2	Sequence 2, Appli
	18	65.8	6.1	9551	1	US-08-056-200-93	Sequence 93, Appl
	19	65.8	6.1	9551	2	US-08-800-644-93	Sequence 93, Appl
	20	63.4	5.9	3489	2	US-08-728-323A-1	Sequence 1, Appli
	21	63.4	5.9	3489	4	US-09-298-568-1	Sequence 1, Appli
	22	63.4	5.9	3489	4	US-09-410-399-1	Sequence 1, Appli
c	23	63.4	5.9	32207	2	US-08-770-379-20	Sequence 20, Appl
c	24	63.4	5.9	32207	3	US-08-757-669A-20	Sequence 20, Appl
c	25	63.4	5.9	32207	4	US-09-230-371A-20	Sequence 20, Appl
	26	60.2	5.6	2793	1	US-08-209-747-1	Sequence 1, Appli
	27	60.2	5.6	2793	1	US-08-458-298-1	Sequence 1, Appli
	28	58.4	5.4	2313	4	US-09-370-838-157	Sequence 157, App
c	29	53.4	4.9	1037	4	US-09-181-585-3	Sequence 3, Appli
c	30	53.4	4.9	1159	4	US-09-181-585-1	Sequence 1, Appli
c	31	53.4	4.9	1471	4	US-09-181-585-2	Sequence 2, Appli
	32	51.8	4.8	1383	4	US-09-252-991A-16519	Sequence 16519, A
c	33	51.8	4.8	2781	4	US-09-252-991A-15980	Sequence 15980, A
	34	51.8	4.8	2952	4	US-09-252-991A-16297	Sequence 16297, A
	35	51.2	4.7	397	3	US-09-253-691-3	Sequence 3, Appli
	36	51	4.7	1995	1	US-08-425-069-3	Sequence 3, Appli
	37	51	4.7	1995	2	US-08-317-844B-3	Sequence 3, Appli
c	38	50.8	4.7	51259	3	US-08-781-891-209	Sequence 209, App
c	39	50.8	4.7	51259	4	US-09-618-166-209	Sequence 209, App
	40	50.2	4.6	1949	4	US-09-740-027-1	Sequence 1, Appli
c	41	49.8	4.6	1455	4	US-09-252-991A-12655	Sequence 12655, A
	42	49.8	4.6	1488	4	US-09-252-991A-13312	Sequence 13312, A
	43	49.6	4.6	3138	4	US-09-434-408-1	Sequence 1, Appli
c	44	49.4	4.6	876	4	US-09-252-991A-8476	Sequence 8476, Ap
	45	49.4	4.6	1776	4	US-09-252-991A-8333	Sequence 8333, Ap

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 6.4%; Score 69; DB 1; Length 7218;

```
  Best Local Similarity   3.1%;  Pred. No. 2.4e-06;
  Matches   12;  Conservative  238;  Mismatches  143;  Indels    0;  Gaps    0;

Qy        386 AGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAGTGGGGACAGCAGCAGAGACAGCTGG 445
Db       1453 AGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1394

Qy        446 GGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTACTACA 505
Db       1393 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1334

Qy        506 AGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGA 565
Db       1333 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1274

Qy        566 TCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCC 625
Db       1273 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1214

Qy        626 TCATGGATGATCACGACCAGGAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGG 685
Db       1213 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1154

Qy        686 CCGCCGCCAAATCCGACGACACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGG 745
Db       1153 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1094

Qy        746 ATGGGGGCAGCGATGGGATGGGAGGGGACGGCA 778
Db       1093 RRRRRRRRRRRRRRRRRRRRRRRRRRRRATCGCA 1061
```

RESULT 2
US-09-249-585A-2
; Sequence 2, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
; FILE REFERENCE: 0867/0D905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
;   LENGTH: 1926
;   TYPE: DNA
;   ORGANISM: Epstein Barr Virus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1926)
;   OTHER INFORMATION: coding strand of EBNA-1 DNA
US-09-249-585A-2

```
  Query Match             6.3%;  Score 68.4;  DB 4;  Length 1926;
  Best Local Similarity  44.0%;  Pred. No. 2.4e-06;
  Matches  288;  Conservative    0;  Mismatches  366;  Indels    0;  Gaps    0;

Qy        406 GGGGGCGAAGAGGAGTGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGA 465
Db        305 GGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 364

Qy        466 CGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAG 525
Db        365 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGG 424

Qy        526 AAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCC 585
Db        425 CAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAG 484

Qy        586 AAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAG 645
Db        485 GGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG 544

Qy        646 GAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGAC 705
Db        545 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGG 604

Qy        706 ACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATG 765
Db        605 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 664

Qy        766 GGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCAC 825
Db        665 GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 724

Qy        826 ACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAG 885
Db        725 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAG 784

Qy        886 AAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTG 945
Db        785 GAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG 844

Qy        946 GGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAG 1005
Db        845 GAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 904

Qy       1006 AACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCG 1059
Db        905 GGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAG 958
```

RESULT 3
US-09-410-399-3
; Sequence 3, Application US/09410399
; Patent No. 6482587
; GENERAL INFORMATION:
; APPLICANT: Robertson, Erle S.
; APPLICANT: Cotter, Murray A.
; TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
; TITLE OF INVENTION: to Genomic Host DNA
; FILE REFERENCE: UM-03778
; CURRENT APPLICATION NUMBER: US/09/410,399
; CURRENT FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 1926
;   TYPE: DNA
;   ORGANISM: Epstein-Barr virus
US-09-410-399-3

```
  Query Match             6.3%;  Score 68.4;  DB 4;  Length 1926;
  Best Local Similarity  44.0%;  Pred. No. 2.4e-06;
  Matches  288;  Conservative    0;  Mismatches  366;  Indels    0;  Gaps    0;

Qy        406 GGGGGCGAAGAGGAGTGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGA 465
Db        305 GGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 364

Qy        466 CGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAG 525
Db        365 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGG 424

Qy        526 AAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCC 585
Db        425 CAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAG 484

Qy        586 AAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAG 645
Db        485 GGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG 544

Qy        646 GAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGAC 705
Db        545 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGG 604
```

```
Qy        706 ACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATG 765
Db        605 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 664

Qy        766 GGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCAC 825
Db        665 GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 724

Qy        826 ACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAG 885
Db        725 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAG 784

Qy        886 AAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTG 945
Db        785 GAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG 844

Qy        946 GGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAG 1005
Db        845 GAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 904

Qy       1006 AACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCG 1059
Db        905 GGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAG 958
```

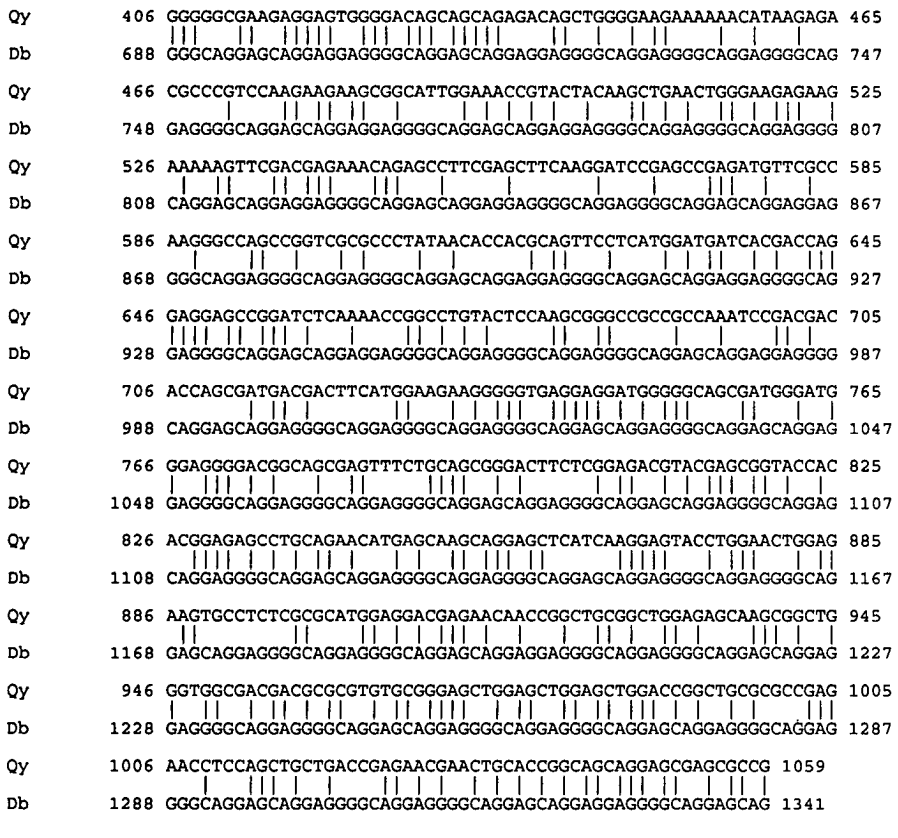
```
RESULT 4
US-09-050-863-2
; Sequence 2, Application US/09050863
; Patent No. 6114111
;  GENERAL INFORMATION:
;    APPLICANT:  Lao, Ying
;    APPLICANT:  Hiang, Betty
;    APPLICANT:  Payan, Don
;    TITLE OF INVENTION:  Mammalian Protein Interaction Cloning
;    TITLE OF INVENTION:  System
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/050,863
;      FILING DATE:  30-MAR-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-65638/DJB/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 949-8711
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2580 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
US-09-050-863-2

  Query Match             6.3%;  Score 68.4;  DB 3;  Length 2580;
  Best Local Similarity   44.0%;  Pred. No. 2.6e-06;
  Matches  288;  Conservative    0;  Mismatches  366;  Indels    0;  Gaps    0;
```

```
Qy        406 GGGGGCGAAGAGGAGTGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGA 465
Db        688 GGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 747

Qy        466 CGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAG 525
Db        748 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGG 807

Qy        526 AAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCC 585
Db        808 CAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAG 867

Qy        586 AAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAG 645
Db        868 GGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG 927

Qy        646 GAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGAC 705
Db        928 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGG 987

Qy        706 ACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATG 765
Db        988 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 1047

Qy        766 GGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCAC 825
Db       1048 GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 1107

Qy        826 ACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAG 885
Db       1108 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAG 1167

Qy        886 AAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTG 945
Db       1168 GAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG 1227

Qy        946 GGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAG 1005
Db       1228 GAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 1287

Qy       1006 AACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCG 1059
Db       1288 GGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAG 1341
```

```
RESULT 5
US-09-359-081-2
; Sequence 2, Application US/09359081
; Patent No. 6316223
;   GENERAL INFORMATION:
;        APPLICANT: Lao, Ying
;                   Hiang, Betty
;                   Payan, Don
;        TITLE OF INVENTION: Mammalian Protein Interaction Cloning
;                            System
;        NUMBER OF SEQUENCES: 5
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;             STREET: 4 Embarcadero Center, Suite 3400
;             CITY: San Francisco
;             STATE: CA
;             COUNTRY: USA
;             ZIP: 94111-4187
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/359,081
;             FILING DATE: 22-Jul-1999
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
```

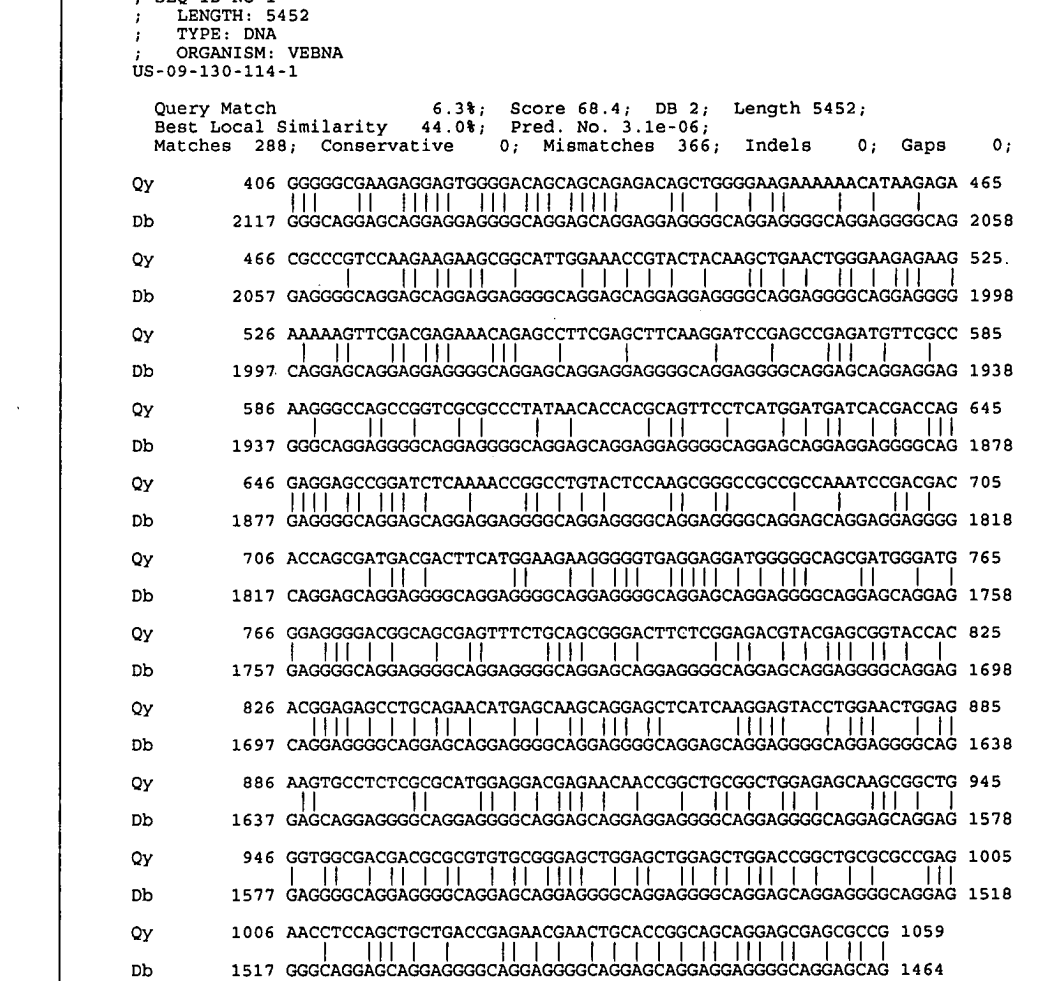
; APPLICATION NUMBER: 09/050,863
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 949-8711
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2580 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-359-081-2

Query Match 6.3%; Score 68.4; DB 4; Length 2580;
Best Local Similarity 44.0%; Pred. No. 2.6e-06;
Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;



RESULT 6
US-09-130-114-1/c
; Sequence 1, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaj, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; TITLE OF INVENTION: From Multiple Transfected Episomes
; FILE REFERENCE: 0867/1D903US1
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 5452
; TYPE: DNA
; ORGANISM: VEBNA
US-09-130-114-1

Query Match 6.3%; Score 68.4; DB 2; Length 5452;
Best Local Similarity 44.0%; Pred. No. 3.1e-06;
Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;



RESULT 7
US-09-647-344A-14/c
; Sequence 14, Application US/09647344A
; Patent No. 6586180
; GENERAL INFORMATION:
; APPLICANT: Ruffner, Duane E.
; APPLICANT: Pierce, Michael L.
; APPLICANT: Chen, Zhidong

```
;  TITLE OF INVENTION: Directed Antisense Libraries
;  FILE REFERENCE: T6678.PCT.US
;  CURRENT APPLICATION NUMBER: US/09/647,344A
;  CURRENT FILING DATE:  2000-12-04
;  PRIOR APPLICATION NUMBER: PCT/US99/06742
;  PRIOR FILING DATE: 1999-03-28
;  NUMBER OF SEQ ID NOS: 50
; SEQ ID NO 14
;    LENGTH: 8705
;    TYPE: DNA
;    ORGANISM: Artificial Sequence
;    FEATURE:
;    OTHER INFORMATION: pShuttle
US-09-647-344A-14

  Query Match             6.3%;  Score 68.4;  DB 4;  Length 8705;
  Best Local Similarity   44.0%;  Pred. No. 3.4e-06;
  Matches  288;  Conservative    0;  Mismatches  366;  Indels    0;  Gaps    0;

Qy        406 GGGGGCGAAGAGGAGTGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGA 465
Db       7982 GGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 7923

Qy        466 CGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAG 525
Db       7922 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGG 7863

Qy        526 AAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCC 585
Db       7862 CAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAG 7803

Qy        586 AAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAG 645
Db       7802 GGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG 7743

Qy        646 GAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGAC 705
Db       7742 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGG 7683

Qy        706 ACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATG 765
Db       7682 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 7623

Qy        766 GGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCAC 825
Db       7622 GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 7563

Qy        826 ACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAG 885
Db       7562 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAG 7503

Qy        886 AAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTG 945
Db       7502 GAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG 7443

Qy        946 GGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAG 1005
Db       7442 GAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 7383

Qy       1006 AACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCG 1059
Db       7382 GGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAG 7329


RESULT 8
US-08-910-647-1
; Sequence 1, Application US/08910647
; Patent No. 6251433
;  GENERAL INFORMATION:
;    APPLICANT:  Zuckermann et al.
;    TITLE OF INVENTION:  Compositions and Methods for
;    TITLE OF INVENTION:  Polynucleotide Delivery
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Chiron Corporation
;      STREET:  4560 Horton Street
;      CITY:  Emeryville
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  94608-2916
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/910,647
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fujita, Sharon M.
;      REGISTRATION NUMBER:  38,459
;      REFERENCE/DOCKET NUMBER:  1218.002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 923-2706
;      TELEFAX:  (510) 655-3542
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9600 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-910-647-1

  Query Match             6.3%;  Score 68.4;  DB 3;  Length 9600;
  Best Local Similarity   44.0%;  Pred. No. 3.5e-06;
  Matches  288;  Conservative    0;  Mismatches  366;  Indels    0;  Gaps    0;

Qy        406 GGGGGCGAAGAGGAGTGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGA 465
Db        734 GGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 793

Qy        466 CGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAG 525
Db        794 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGG 853

Qy        526 AAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCC 585
Db        854 CAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAG 913

Qy        586 AAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAG 645
Db        914 GGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG 973

Qy        646 GAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGAC 705
Db        974 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGG 1033

Qy        706 ACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATG 765
Db       1034 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 1093

Qy        766 GGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCAC 825
Db       1094 GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 1153

Qy        826 ACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAG 885
Db       1154 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAG 1213

Qy        886 AAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTG 945
Db       1214 GAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG 1273

Qy        946 GGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAG 1005
```

```
Db       1274 GAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 1333

Qy       1006 AACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCG 1059

Db       1334 GGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAG 1387


RESULT 9
US-09-620-925-1
; Sequence 1, Application US/09620925
; Patent No. 6468986
;   GENERAL INFORMATION:
;        APPLICANT: Zuckermann et al.
;        TITLE OF INVENTION: Compositions and Methods for
;                            Polynucleotide Delivery
;        NUMBER OF SEQUENCES: 4
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Chiron Corporation
;             STREET: 4560 Horton Street
;             CITY: Emeryville
;             STATE: California
;             COUNTRY: U.S.A.
;             ZIP: 94608-2916
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/620,925
;             FILING DATE: 21-Jul-2000
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/910,647
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Fujita, Sharon M.
;             REGISTRATION NUMBER: 38,459
;             REFERENCE/DOCKET NUMBER: 1218.002
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (510) 923-2706
;             TELEFAX: (510) 655-3542
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 9600 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-620-925-1

  Query Match             6.3%;  Score 68.4;  DB 4;  Length 9600;
  Best Local Similarity   44.0%;  Pred. No. 3.5e-06;
  Matches  288;  Conservative    0;  Mismatches  366;  Indels    0;  Gaps    0;

Qy        406 GGGGGCGAAGAGGAGTGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGA 465

Db        734 GGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 793

Qy        466 CGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAG 525

Db        794 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGG 853

Qy        526 AAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCC 585

Db        854 CAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAG 913

Qy        586 AAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAG 645

Db        914 GGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG 973

Qy        646 GAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGAC 705

Db        974 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGG 1033

Qy        706 ACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATG 765

Db       1034 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 1093

Qy        766 GGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCAC 825

Db       1094 GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 1153

Qy        826 ACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAG 885

Db       1154 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAG 1213

Qy        886 AAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTG 945

Db       1214 GAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG 1273

Qy        946 GGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAG 1005

Db       1274 GAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 1333

Qy       1006 AACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCG 1059

Db       1334 GGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAG 1387


RESULT 10
US-07-884-811-15
; Sequence 15, Application US/07884811
; Patent No. 5316921
;  GENERAL INFORMATION:
;    APPLICANT:  Godowski, Paul J.  Lokker, Nathalie A.  Mark, Melanie R.
;    TITLE OF INVENTION:  SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/884,811
;      FILING DATE:  19920518
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Ginger R.
;      REGISTRATION NUMBER:  33,055
;      REFERENCE/DOCKET NUMBER:  755.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-3216
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 bases
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-07-884-811-15

  Query Match             6.3%;  Score 68.4;  DB 1;  Length 10596;
```

Best Local Similarity 44.0%; Pred. No. 3.6e-06;
Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

```
Qy        406 GGGGGCGAAGAGGAGTGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGA 465
              |||   ||  |||||  ||| ||| |||||    ||  |  | ||    |  |   |
Db       2268 GGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 2327

Qy        466 CGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAG 525
                  |    || || ||  |     | | | | |  |   || | |  || | ||| |
Db       2328 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGG 2387

Qy        526 AAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCC 585
               |  ||   || |||   |||  |      |      |    |    |||  |   |
Db       2388 CAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAG 2447

Qy        586 AAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAG 645
               |    ||  |   | |    | |      | ||    |    | | ||  | |  |||
Db       2448 GGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG 2507

Qy        646 GAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGAC 705
              |||| || ||| |    |    || | | |     ||  ||      |   |    || |
Db       2508 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 2567

Qy        706 ACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATG 765
                 | || |      ||   | | |||  ||||| | | |||    ||   |  |
Db       2568 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 2627

Qy        766 GGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCAC 825
              |  ||| | |   |  ||     ||||  | |      | ||  | | |||  | |  |
Db       2628 GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 2687

Qy        826 ACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAG 885
               |||| | | | ||  |  | |  || ||| ||       |||||   | |||   | ||
Db       2688 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAG 2747

Qy        886 AAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTG 945
               ||      ||   || | | ||| |  |   |  || | ||  |     |||  | |
Db       2748 GAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 2807

Qy        946 GGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAG 1005
              |  ||  | || | ||  | || |||| | ||  || || ||| | |  | |    |||
Db       2808 GAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 2867

Qy       1006 AACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCG 1059
                 |   ||| |  |    || | |  | | | | | || ||| ||  | || |
Db       2868 GGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAG 2921
```

RESULT 11
US-07-885-971-15
; Sequence 15, Application US/07885971
; Patent No. 5328837
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/885,971
; FILING DATE: 19920518
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
US-07-885-971-15

Query Match 6.3%; Score 68.4; DB 1; Length 10596;
Best Local Similarity 44.0%; Pred. No. 3.6e-06;
Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

```
Qy        406 GGGGGCGAAGAGGAGTGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGA 465
              |||   ||  |||||  ||| ||| |||||    ||  |  | ||    |  |   |
Db       2268 GGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 2327

Qy        466 CGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAG 525
                  |    || || ||  |     | | | | |  |   || | |  || | ||| |
Db       2328 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGG 2387

Qy        526 AAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCC 585
               |  ||   || |||   |||  |      |      |    |    |||  |   |
Db       2388 CAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAG 2447

Qy        586 AAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAG 645
               |    ||  |   | |    | |      | ||    |    | | ||  | |  |||
Db       2448 GGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG 2507

Qy        646 GAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGAC 705
              |||| || ||| |    |    || | | |     ||  ||      |   |    || |
Db       2508 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGG 2567

Qy        706 ACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATG 765
                 | || |      ||   | | |||  ||||| | | |||    ||   |  |
Db       2568 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 2627

Qy        766 GGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCAC 825
              |  ||| | |   |  ||     ||||  | |      | ||  | | |||  | |  |
Db       2628 GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 2687

Qy        826 ACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAG 885
               |||| | | | ||  |  | |  || ||| ||       |||||   | |||   | ||
Db       2688 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAG 2747

Qy        886 AAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTG 945
               ||      ||   || | | ||| |  |   |  || | ||  |     |||  | |
Db       2748 GAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG 2807

Qy        946 GGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAG 1005
              |  ||  | || | ||  | || |||| | ||  || || ||| | |  | |    |||
Db       2808 GAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 2867

Qy       1006 AACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCG 1059
                 |   ||| |  |    || | |  | | | | | || ||| ||  | || |
Db       2868 GGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAG 2921
```

RESULT 12
US-08-087-783A-15
; Sequence 15, Application US/08087783A
; Patent No. 5547856
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.

```
;    TITLE OF INVENTION:  HEPATOCYTE GROWTH FACTOR VARIANTS
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/087,783A
;      FILING DATE:  13-Jul-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/884811
;      FILING DATE:  18-MAY-92
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/885971
;      FILING DATE:  18-MAY-92
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Marschang, Diane L.
;      REGISTRATION NUMBER:  35,600
;      REFERENCE/DOCKET NUMBER:  P0755779P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-5416
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-087-783A-15

  Query Match             6.3%;  Score 68.4;  DB 1;  Length 10596;
  Best Local Similarity   44.0%;  Pred. No. 3.6e-06;
  Matches  288;  Conservative    0;  Mismatches  366;  Indels    0;  Gaps    0;

Qy        406 GGGGGCGAAGAGGAGTGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGA 465
              |||   ||  |||||  ||| ||| |||||    ||  |  | ||    |  |   |
Db       2268 GGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 2327

Qy        466 CGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAG 525
                |    || || ||  |     | | | | |   || | |  || | ||| |
Db       2328 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGG 2387

Qy        526 AAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCC 585
               |  ||   || |||   |||   |       |       |    |   ||| |   |
Db       2388 CAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAG 2447

Qy        586 AAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAG 645
               |    ||  |   | |   |  |      | ||   |   | | ||  | |  |||
Db       2448 GGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG 2507

Qy        646 GAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGAC 705
              |||| || ||| |   |    || | | |     ||  ||      |   |    || |
Db       2508 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGG 2567

Qy        706 ACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATG 765
                | || |        ||   | | |||  ||||| | | |||     ||   |  |
Db       2568 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 2627

Qy        766 GGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCAC 825
               |  ||| | |   |  ||      ||||  | |       | ||  | | |||  || |  |
Db       2628 GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 2687

Qy        826 ACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAG 885
```

```
              |||| | | | || |   | |  || ||| ||      |||||   | |||   | ||
Db       2688 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAG 2747

Qy        886 AAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTG 945
               ||      ||   || | | ||| | . |   |  || |  || |   ||| |  |
Db       2748 GAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG 2807

Qy        946 GGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAG 1005
               |  ||  | || | ||  | || ||||  | ||  || || ||| | |  | |    |||
Db       2808 GAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 2867

Qy       1006 AACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCG 1059
                  |  ||| |  |    || | | | | | | | || ||| ||  | || |
Db       2868 GGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAG 2921


RESULT 13
US-08-194-088B-15
; Sequence 15, Application US/08194088B
; Patent No. 5580963
;  GENERAL INFORMATION:
;    APPLICANT:  Godowski, Paul J.  Lokker, Nathalie A.  Mark, Melanie R.
;    TITLE OF INVENTION:  SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/194,088B
;      FILING DATE:  09-FEB-1994
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/884811
;      FILING DATE:  18-MAY-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gallegos, R. Thomas
;      REGISTRATION NUMBER:  32,692
;      REFERENCE/DOCKET NUMBER:  755D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-2614
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-194-088B-15

  Query Match             6.3%;  Score 68.4;  DB 1;  Length 10596;
  Best Local Similarity   44.0%;  Pred. No. 3.6e-06;
  Matches  288;  Conservative    0;  Mismatches  366;  Indels    0;  Gaps    0;

Qy        406 GGGGGCGAAGAGGAGTGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGA 465
              |||   ||  |||||  ||| ||| |||||    ||  |  | ||    |  |   |
Db       2268 GGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 2327

Qy        466 CGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAG 525
                |    || || ||  |     | | | | |   || | |  || | ||| |
Db       2328 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGG 2387

Qy        526 AAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCC 585
               |  ||   || |||   |||   |       |       |    |   ||| |   |
```

```
Db       2388 CAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAG 2447

Qy        586 AAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAG 645

Db       2448 GGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG 2507

Qy        646 GAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGAC 705

Db       2508 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGG 2567

Qy        706 ACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATG 765

Db       2568 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 2627

Qy        766 GGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCAC 825

Db       2628 GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 2687

Qy        826 ACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAG 885

Db       2688 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAG 2747

Qy        886 AAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTG 945

Db       2748 GAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG 2807

Qy        946 GGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAG 1005

Db       2808 GAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 2867

Qy       1006 AACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCG 1059

Db       2868 GGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAG 2921


RESULT 14
US-08-194-087-15
; Sequence 15, Application US/08194087
; Patent No. 5879910
;  GENERAL INFORMATION:
;    APPLICANT:  Godowski, Paul J.  Lokker, Nathalie A.  Mark, Melanie R.
;    TITLE OF INVENTION:  HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/194,087
;      FILING DATE:  18-MAY-1992
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Ginger R.
;      REGISTRATION NUMBER:  33,055
;      REFERENCE/DOCKET NUMBER:  779
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-3216
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-194-087-15

  Query Match             6.3%;  Score 68.4;  DB 2;  Length 10596;
  Best Local Similarity   44.0%;  Pred. No. 3.6e-06;
  Matches  288;  Conservative    0;  Mismatches  366;  Indels    0;  Gaps    0;

Qy        406 GGGGGCGAAGAGGAGTGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGA 465

Db       2268 GGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 2327

Qy        466 CGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAG 525

Db       2328 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGG 2387

Qy        526 AAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCC 585

Db       2388 CAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAG 2447

Qy        586 AAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAG 645

Db       2448 GGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG 2507

Qy        646 GAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGAC 705

Db       2508 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGG 2567

Qy        706 ACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATG 765

Db       2568 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 2627

Qy        766 GGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCAC 825

Db       2628 GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 2687

Qy        826 ACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAG 885

Db       2688 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAG 2747

Qy        886 AAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTG 945

Db       2748 GAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG 2807

Qy        946 GGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAG 1005

Db       2808 GAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 2867

Qy       1006 AACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCG 1059

Db       2868 GGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAG 2921


RESULT 15
PCT-US93-04648-15
; Sequence 15, Application PC/TUS9304648
;  GENERAL INFORMATION:
;    APPLICANT:  Genentech, Inc., Godowski, Paul J., Lokker, Natalie A., Mark, Melanie R.
;    TITLE OF INVENTION:  HEPATOCYTE GROWTH FACTOR VARIANTS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/04648
;      FILING DATE:  19930517
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/884811
;      FILING DATE:  18-MAY-92
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/885971
;      FILING DATE:  18-MAY-92
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Ginger R.
;      REGISTRATION NUMBER:  33,055
;      REFERENCE/DOCKET NUMBER:  755,779P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-3216
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 bases
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
PCT-US93-04648-15

  Query Match             6.3%;  Score 68.4;  DB 5;  Length 10596;
  Best Local Similarity   44.0%;  Pred. No. 3.6e-06;
  Matches  288;  Conservative    0;  Mismatches  366;  Indels    0;  Gaps    0;

Qy        406 GGGGGCGAAGAGGAGTGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGA 465
              |||    ||  |||||  ||| ||| |||||    ||  |  | ||    |  |  |
Db       2268 GGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 2327

Qy        466 CGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAG 525
                |    || || ||  |     | | | | |  |   || | |  || | |||  |
Db       2328 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGG 2387

Qy        526 AAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCC 585
              |  ||   || |||   |||   |      |       |    |    |||  |   |
Db       2388 CAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAG 2447

Qy        586 AAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAG 645
              |   ||  |  | |   | |   | | ||   |   | | ||  | |  |||
Db       2448 GGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG 2507

Qy        646 GAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGAC 705
              |||| || ||| |   |    || | | |      ||  ||    |   |    || |
Db       2508 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGG 2567

Qy        706 ACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATG 765
                | || |         ||   | | |||  ||||| | | |||    ||   |  |
Db       2568 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 2627

Qy        766 GGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCAC 825
              |  ||| | |   |  ||     ||||  | |      | ||  | | ||| || |  |
Db       2628 GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 2687

Qy        826 ACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAG 885
              ||||  | | | || |   | |  || ||| ||      |||||   | |||   | ||
Db       2688 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAG 2747

Qy        886 AAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTG 945
               ||          ||    || | | ||| |  |   |  || |  || |     ||| |  |
Db       2748 GAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG 2807

Qy        946 GGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAG 1005
              |  ||  | || | ||  | || ||||  | ||  || || ||| | |  | |   |||
Db       2808 GAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 2867

Qy       1006 AACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCG 1059
                |   ||| |  |    || | |  | | | | | || ||| ||  | || |
Db       2868 GGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAG 2921
```

Search completed: February 5, 2004, 12:05:27
Job time : 99 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 5, 2004, 11:18:43 ; Search time 486 Seconds
(without alignments)
8185.851 Million cell updates/sec

Title: US-09-972-758A-1
Perfect score: 1080
Sequence: 1 atggccgagccattcttgtc.....tttccaagtttggagactag 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1080	100.0	1080	10	US-09-972-758-1	Sequence 1, Appli
	2	1076.8	99.7	2199	9	US-09-745-763-33	Sequence 33, Appl
	3	1076.8	99.7	3624	15	US-10-205-823-171	Sequence 171, App
c	4	434.2	40.2	1083	12	US-10-264-237-629	Sequence 629, App
c	5	368.8	34.1	495	11	US-09-918-995-403	Sequence 403, App
	6	362.8	33.6	414	10	US-09-983-965-4895	Sequence 4895, Ap
	7	355.2	32.9	461	10	US-09-954-531-813	Sequence 813, App
	8	174.8	16.2	1330	13	US-10-094-749-1223	Sequence 1223, Ap
c	9	173.2	16.0	2048	12	US-10-108-260A-1496	Sequence 1496, Ap
	10	68.4	6.3	1926	13	US-10-294-804-3	Sequence 3, Appli
c	11	68.4	6.3	8705	15	US-10-291-230-14	Sequence 14, Appl
c	12	68.4	6.3	8705	15	US-10-291-249-14	Sequence 14, Appl
	13	68.4	6.3	9600	13	US-10-278-751-1	Sequence 1, Appli
c	14	68.4	6.3	10233	13	US-10-050-898-283	Sequence 283, App
c	15	68.4	6.3	10285	13	US-10-050-902-283	Sequence 283, App
	16	67.8	6.3	1548	13	US-10-322-774-5	Sequence 5, Appli
	17	67.8	6.3	3194	13	US-09-814-353-20004	Sequence 20004, A
	18	67.8	6.3	4270	10	US-09-880-107-2427	Sequence 2427, Ap
	19	67.8	6.3	4270	10	US-09-954-531-1352	Sequence 1352, Ap
	20	67.8	6.3	4653	13	US-10-269-909-45	Sequence 45, Appl
	21	63.4	5.9	3489	13	US-10-294-804-1	Sequence 1, Appli
	22	60.8	5.6	2108	10	US-09-962-832-225	Sequence 225, App
	23	58.4	5.4	2313	10	US-09-738-973-157	Sequence 157, App
	24	58.4	5.4	2313	10	US-09-854-133-157	Sequence 157, App
	25	58.4	5.4	2313	15	US-10-144-649A-157	Sequence 157, App
	26	58.4	5.4	2314	10	US-09-764-868-12	Sequence 12, Appl
	27	58.4	5.4	2446	12	US-10-057-475B-10970	Sequence 10970, A
	28	58.4	5.4	2446	12	US-10-154-884B-10970	Sequence 10970, A
	29	58.4	5.4	2802	13	US-10-233-045-35	Sequence 35, Appl
	30	53.6	5.0	607	15	US-10-043-487-122	Sequence 122, App
	31	53.6	5.0	1609	15	US-10-043-487-105	Sequence 105, App
	32	53.6	5.0	3705	13	US-10-116-275-315	Sequence 315, App
c	33	53.4	4.9	1037	12	US-10-373-667-3	Sequence 3, Appli
c	34	53.4	4.9	1159	12	US-10-373-667-1	Sequence 1, Appli
c	35	53.4	4.9	1471	12	US-10-373-667-2	Sequence 2, Appli
	36	53.2	4.9	628	13	US-10-029-386-22859	Sequence 22859, A
	37	52.2	4.8	3243	15	US-10-156-761-2685	Sequence 2685, Ap
	38	52.2	4.8	9025608	15	US-10-156-761-1	Sequence 1, Appli
	39	52	4.8	225883	15	US-10-175-523-57	Sequence 57, Appl
	40	51.8	4.8	2529	12	US-10-369-493-31518	Sequence 31518, A
	41	51.4	4.8	2479	12	US-10-108-260A-1966	Sequence 1966, Ap
c	42	51.4	4.8	1223197	13	US-10-027-632-179264	Sequence 179264,
c	43	51.4	4.8	1223197	14	US-10-027-632-179264	Sequence 179264,
	44	51.2	4.7	1799	12	US-10-108-260A-1885	Sequence 1885, Ap
	45	50.8	4.7	2183	12	US-10-104-047-1064	Sequence 1064, Ap

ALIGNMENTS

RESULT 1
US-09-972-758-1
; Sequence 1, Application US/09972758
; Patent No. US20020160497A1
; GENERAL INFORMATION:
; APPLICANT: Case Western Reserve University
; APPLICANT: Montano, Monica
; APPLICANT: Wittman, Bryan
; TITLE OF INVENTION: Suppressors of Human Breast Cancer Cell Growth
; FILE REFERENCE: 27708/04004
; CURRENT APPLICATION NUMBER: US/09/972,758
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: US 60/238,187
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1080
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-972-758-1

Query Match 100.0%; Score 1080; DB 10; Length 1080;
Best Local Similarity 100.0%; Pred. No. 4.7e-269;
Matches 1080; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60

Qy         61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120

Qy        121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180
```

```
Qy       181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240
Db       181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240

Qy       241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300
Db       241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300

Qy       301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360
Db       301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360

Qy       361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 420
Db       361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 420

Qy       421 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAG 480
Db       421 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAG 480

Qy       481 AAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAG 540
Db       481 AAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAG 540

Qy       541 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 600
Db       541 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 600

Qy       601 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 660
Db       601 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 660

Qy       661 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 720
Db       661 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 720

Qy       721 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 780
Db       721 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 780

Qy       781 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 840
Db       781 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 840

Qy       841 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 900
Db       841 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 900

Qy       901 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 960
Db       901 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 960

Qy       961 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1020
Db       961 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1020

Qy      1021 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080
Db      1021 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080
```

```
RESULT 2
US-09-745-763-33
; Sequence 33, Application US/09745763
; Patent No. US20020065394A1
;   GENERAL INFORMATION:
;        APPLICANT: Jacobs, Kenneth
;                   McCoy, John M.
;                   LaVallie, Edward R.
;                   Collins-Racie, Lisa A.
;                   Evans, Cheryl
;                   Merberg, David
;                   Treacy, Maurice
;                   Spaulding, Vikki
;        TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
;                            ENCODING THEM
;        NUMBER OF SEQUENCES: 219
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genetics Institute, Inc.
;             STREET: 87 CambridgePark Drive
;             CITY: Cambridge
;             STATE: MA
;             COUNTRY: U.S.A.
;             ZIP: 02140
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/745,763
;             FILING DATE: 18-Jun-2000
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Sprunger, Suzanne A.
;             REGISTRATION NUMBER: 41,323
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (617) 498-8284
;             TELEFAX: (617) 876-5851
;   INFORMATION FOR SEQ ID NO: 33:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 2199 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: <Unknown>
US-09-745-763-33

  Query Match             99.7%;  Score 1076.8;  DB 9;  Length 2199;
  Best Local Similarity   99.8%;  Pred. No. 3.6e-268;
  Matches 1078;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60
Db       707 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 766

Qy        61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120
Db       767 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 826

Qy       121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180
Db       827 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 886

Qy       181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240
Db       887 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 946

Qy       241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300
Db       947 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 1006

Qy       301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360
Db      1007 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 1066

Qy       361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 420
Db      1067 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 1126

Qy       421 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAG 480
Db      1127 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 1186

Qy       481 AAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAG 540
```

```
Db       1187 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 1246

Qy        541 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 600
Db       1247 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 1306

Qy        601 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 660
Db       1307 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 1366

Qy        661 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 720
Db       1367 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 1426

Qy        721 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 780
Db       1427 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 1486

Qy        781 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 840
Db       1487 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 1546

Qy        841 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 900
Db       1547 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 1606

Qy        901 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 960
Db       1607 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 1666

Qy        961 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1020
Db       1667 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1726

Qy       1021 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080
Db       1727 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1786
```

RESULT 3
US-10-205-823-171
; Sequence 171, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbatcheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Wonsey, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 171
; LENGTH: 3624
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-205-823-171

Query Match 99.7%; Score 1076.8; DB 15; Length 3624;
Best Local Similarity 99.8%; Pred. No. 3.9e-268;
Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

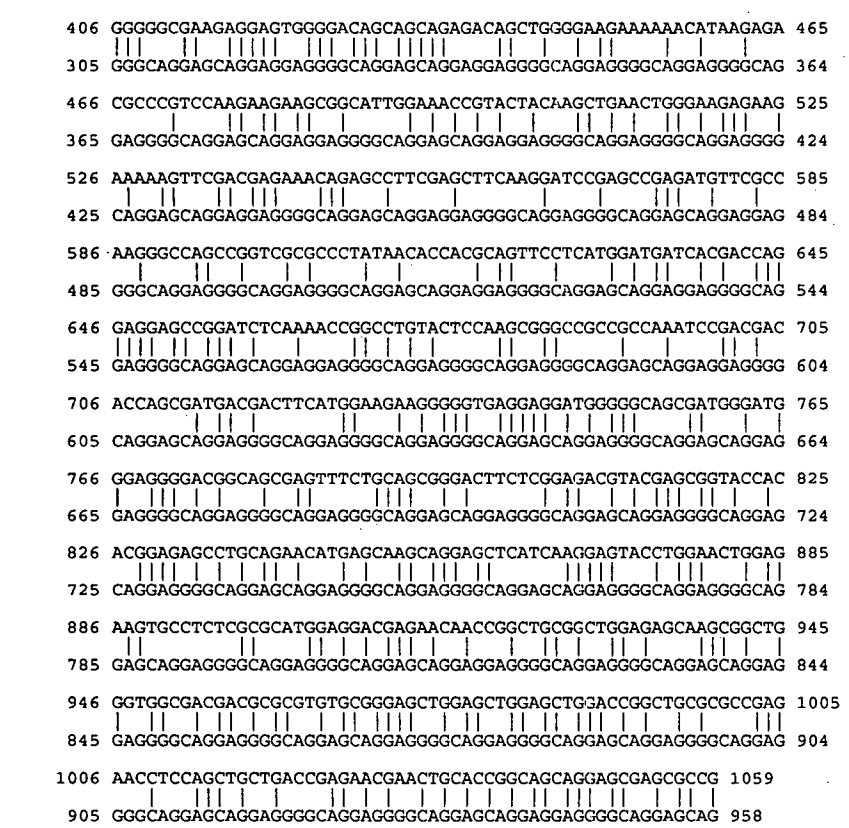
```
Qy           1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60
Db         690 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 749

Qy          61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120
Db         750 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 809

Qy         121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180
Db         810 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 869

Qy         181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240
Db         870 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 929

Qy         241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300
Db         930 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 989

Qy         301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360
Db         990 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 1049

Qy         361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 420
Db        1050 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 1109

Qy         421 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAG 480
Db        1110 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 1169

Qy         481 AAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAG 540
Db        1170 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 1229

Qy         541 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 600
Db        1230 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 1289

Qy         601 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 660
Db        1290 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 1349

Qy         661 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 720
Db        1350 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 1409

Qy         721 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 780
Db        1410 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 1469

Qy         781 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 840
Db        1470 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 1529

Qy         841 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 900
Db        1530 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 1589

Qy         901 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 960
```

```
Db       1590 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 1649

Qy        961 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1650 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1709

Qy       1021 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1710 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGACTAG 1769


RESULT 4
US-10-264-237-629/c
; Sequence 629, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
;  APPLICANT: Birse et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA131P1
;  CURRENT APPLICATION NUMBER: US/10/264,237
;  CURRENT FILING DATE:  2002-10-04
;  PRIOR APPLICATION NUMBER: PCT/US01/16450
;  PRIOR FILING DATE: 2001-05-18
;  PRIOR APPLICATION NUMBER: US 60/205,515
;  PRIOR FILING DATE: 2000-05-19
;  NUMBER OF SEQ ID NOS: 2876
;  SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 629
;   LENGTH: 1083
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (16)..(16)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (27)..(27)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (32)..(32)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (111)..(111)
;   OTHER INFORMATION: n equals a,t,g, or c
US-10-264-237-629

  Query Match             40.2%;  Score 434.2;  DB 12;  Length 1083;
  Best Local Similarity   98.9%;  Pred. No. 2.6e-102;
  Matches  444;  Conservative    3;  Mismatches    1;  Indels    1;  Gaps    1;

Qy        632 ATGATCACGACCAGGAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCCG 691
              |||||||||||||||||||||||| ||||||||||||:||||||||:|||:||||||
Db       1083 ATGATCACGACCAGGAGGAGCCGGACCTCAAAACCGGCCTKTACTCCAARCGGSCCGCCG 1024

Qy        692 CCAAATCCGACGACACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGG 751
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1023 -CAAATCCGACGACACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGG 965

Qy        752 GCAGCGATGGGATGGGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGT 811
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        964 GCAGCGATGGGATGGGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGT 905

Qy        812 ACGAGCGGTACCACACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGT 871
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        904 ACGAGCGGTACCACACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGT 845

Qy        872 ACCTGGAACTGGAGAAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGG 931
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        844 ACCTGGAACTGGAGAAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGG 785

Qy        932 AGAGCAAGCGGCTGGGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACC 991
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        784 AGAGCAAGCGGCTGGGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACC 725

Qy        992 GGCTGCGCGCCGAGAACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGC 1051
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 GGCTGCGCGCCGAGAACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGC 665

Qy       1052 GAGCGCCGCTTTCCAAGTTTGGAGACTAG 1080
              |||||||||||||||||||||||||||||
Db        664 GAGCGCCGCTTTCCAAGTTTGGAGACTAG 636


RESULT 5
US-09-918-995-403/c
; Sequence 403, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 403
;   LENGTH: 495
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(495)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-403

  Query Match             34.1%;  Score 368.8;  DB 11;  Length 495;
  Best Local Similarity   99.5%;  Pred. No. 1.7e-85;
  Matches  370;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        709 AGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGA 768
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        495 AGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGCATGGGGGCAGCGATGGGATGGGA 436

Qy        769 GGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACG 828
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        435 GGGCACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACG 376

Qy        829 GAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAG 888
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        375 GAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAG 316

Qy        889 TGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGT 948
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        315 TGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGT 256

Qy        949 GGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAAC 1008
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        255 GGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAAC 196

Qy       1009 CTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAG 1068
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         195 CTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAG 136

Qy       1069 TTTGGAGACTAG 1080
              ||||||||||||
Db         135 TTTGGAGACTAG 124


RESULT 6
US-09-983-965-4895
```

```
; Sequence 4895, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 37-21(10297)C
;  CURRENT APPLICATION NUMBER: US/09/983,965
;  CURRENT FILING DATE:  2001-10-26
;  PRIOR APPLICATION NUMBER: US 09/465,231
;  PRIOR FILING DATE: 1999-12-15
;  PRIOR APPLICATION NUMBER: US 60/113,678
;  PRIOR FILING DATE: 1998-12-17
;  NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 4895
;   LENGTH: 414
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 24-LIB34-008-Q1-E1-F7
US-09-983-965-4895

  Query Match             33.6%;  Score 362.8;  DB 10;  Length 414;
  Best Local Similarity   92.3%;  Pred. No. 6e-84;
  Matches  382;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy        462 GAGACGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGA 521
              ||||||||| |||||||||||||||| |||||||||||||| | ||||| |||||| ||
Db          1 GAGACGCCCCTCCAAGAAGAAGCGGCTTTGGAAACCGTACTATACGCTGACCTGGGAGGA 60

Qy        522 GAAGAAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTT 581
              |||||||||||||| || |||||||||| |||||||| ||||| ||||||||||||||
Db         61 GAAGAAAAAGTTCGATGAAAAACAGAGCCTGCGAGCTTCGAGGATTCGAGCCGAGATGTT 120

Qy        582 CGCCAAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGA 641
              ||||||||||||||| || || |||||||||||||||||||||||||||||||| |||||
Db        121 CGCCAAGGGCCAGCCAGTTGCTCCCTATAACACCACGCAGTTCCTCATGGATGACCACGA 180

Qy        642 CCAGGAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGA 701
              |||||||||||||||| |||||||||| ||  |||| |||||||| ||||||||||
Db        181 CCAGGAGGAGCCGGATCTTAAAACCGGCCTCTATCCCAAACGGGCCGCTGCCAAATCCGA 240

Qy        702 CGACACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGG 761
              ||||||||||||||| ||||| |||||||||| ||| ||||||||||||||||||| ||
Db        241 CGACACCAGCGATGAGGACTTTATGGAAGAAGCGGGCGAGGAGGATGGGGGCAGCGACGG 300

Qy        762 GATGGGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTA 821
              ||||||||| |||||||||||||||||||||||||||||||||||  || ||||||||
Db        301 GATGGGAGGAGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGAACTATGAGCGGTA 360

Qy        822 CCACACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTACCT 875
              |||| |||||||||||||||||||||||||||||||||||||||| |||||||
Db        361 CCACGCGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAAGAGTACCT 414


RESULT 7
US-09-954-531-813
; Sequence 813, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
;  APPLICANT: Weaver, Zoe
;  TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
;  TITLE OF INVENTION:  Gene Sets
;  FILE REFERENCE: 689290-77
;  CURRENT APPLICATION NUMBER: US/09/954,531
;  CURRENT FILING DATE:  2002-05-02
;  PRIOR APPLICATION NUMBER: US/60/233,133
;  PRIOR FILING DATE: 2000-09-18
;  PRIOR APPLICATION NUMBER: US/60/234,009
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,034
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,509
;  PRIOR FILING DATE: 2000-09-22
;  PRIOR APPLICATION NUMBER: US/60/234,567
;  PRIOR FILING DATE: 2000-09-22
;  NUMBER OF SEQ ID NOS: 1392
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 813
;   LENGTH: 461
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(461)
;   OTHER INFORMATION: n=a,t,g or c
US-09-954-531-813

  Query Match             32.9%;  Score 355.2;  DB 10;  Length 461;
  Best Local Similarity   95.7%;  Pred. No. 5.6e-82;
  Matches  374;  Conservative    0;  Mismatches   16;  Indels    1;  Gaps    1;

Qy          1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         72 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 131

Qy         61 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 120
              |||||||||||||||||||||||||||||||||||||||||||||     ||||||||||
Db        132 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGNCACGGAGCGGGTG 191

Qy        121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 251

Qy        181 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 240
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 CCGGACGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 311

Qy        241 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        312 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 371

Qy        301 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 360
              |||||||||| | | |    ||||||| ||||||||||||||||||||| | ||||
Db        372 -GCGACTTCCCGGCNCGGCAGAAAGTGGAGCCGACGCCCGAGGCCGAGCTGCTTGNCCAG 430

Qy        361 CCTTGTCATGACTCCGAGGCCAGTAAGTTGG 391
              |||||||||||||||||| ||||||||||
Db        431 CCTTGTCATGACTCCGAGGGCAGTAAGTTGG 461


RESULT 8
US-10-094-749-1223
; Sequence 1223, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
;  APPLICANT: ISOGAI, TAKAO
;  APPLICANT:  SUGIYAMA, TOMOYASU
;  APPLICANT:  OTSUKI, TETSUJI
;  APPLICANT:  WAKAMATSU, AI
;  APPLICANT:  SATO, HIROYUKI
;  APPLICANT:  ISHII, SHIZUKO
;  APPLICANT:  YAMAMOTO, JUN-ICHI
;  APPLICANT:  ISONO, YUUKO
;  APPLICANT:  HIO, YURI
;  APPLICANT:  OTSUKA, KAORU
;  APPLICANT:  NAGAI, KEIICHI
;  APPLICANT:  IRIE, RYOTARO
;  APPLICANT:  TAMECHIKA, ICHIRO
;  APPLICANT:  SEKI, NAOHIKO
;  APPLICANT:  YOSHIKAWA, TSUTOMU
;  APPLICANT:  OTSUKA, MOTOYUKI
;  APPLICANT:  NAGAHARI, KENJI
```

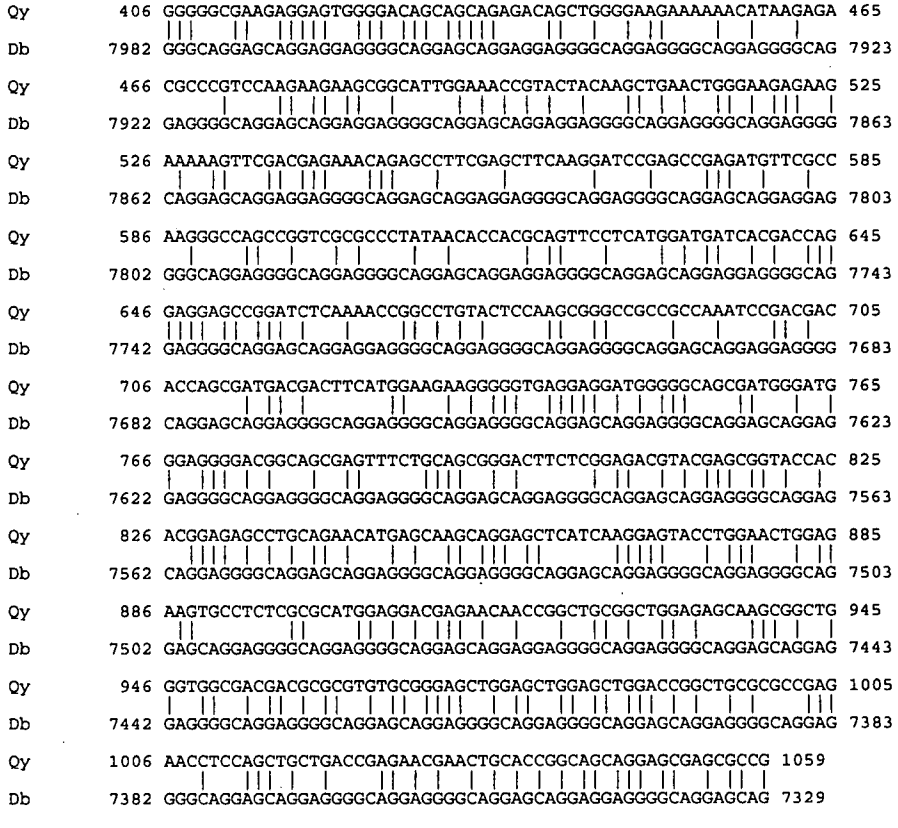
```
;  APPLICANT:  MASUHO, YASUHIKO
;  TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
;  FILE REFERENCE: 084335/0160
;  CURRENT APPLICATION NUMBER: US/10/094,749
;  CURRENT FILING DATE:  2002-03-12
;  PRIOR APPLICATION NUMBER: 60/350,435
;  PRIOR FILING DATE: 2002-01-24
;  PRIOR APPLICATION NUMBER: JP 2001-328381
;  PRIOR FILING DATE: 2001-09-14
;  NUMBER OF SEQ ID NOS: 3381
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1223
;   LENGTH: 1330
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-094-749-1223

  Query Match            16.2%;  Score 174.8;  DB 13;  Length 1330;
  Best Local Similarity   60.5%;  Pred. No. 2.6e-35;
  Matches  362;  Conservative    0;  Mismatches  197;  Indels   39;  Gaps    3;

Qy        441 GCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTA 500
              |||||    ||| |||||      | || || || |||    ||  |||| |||  ||| ||
Db        494 GCTGGCCCGGAAGAAACACCGTCGGCGGCCATCGAAGCGCAAAAGGCACTGGCGACCCTA 553

Qy        501 CTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTC 560
              |    |||||| |||||  |||||  || ||   || ||||  |||||||   | || ||
Db        554 CCTGGAGCTGAGCTGGGCTGAGAAACAACAGCGGGATGAGAGGCAGAGCCAGAGGGCCTC 613

Qy        561 AAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTCGCGCCCTATAACACCACGCA 620
                || |||| |  |||||||||||||| |||||||| || || ||||| ||||||| ||
Db        614 CCGGGTCCGCGAAGAGATGTTCGCCAAAGGCCAGCCCGTGGCCCCCTACAACACCACCCA 673

Qy        621 GTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAA 680
              |||||| ||| ||||    |||| |||||||||          |||  |  ||| | |||
Db        674 GTTCCTGATGAATGACAGGGACCCGGAGGAGCC--------CAACTTGGATGTGCCCCAT 725

Qy        681 GCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGACTTCATGGAAGAAGGGGGTGA 740
              | |  |  || || |     |                           ||||
Db        726 GGGATCTCCCACCCAGGTTCCAG---------------------------TGGGGAGAG 757

Qy        741 GGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTT 800
              ||||  |||| ||| ||||||  |||  | |    | | ||||| | || |  ||||||
Db        758 TGAGGCCGGGGACAGTGATGGGCGGGGCCGAGCGCACGGTGAGTTCCAGCGGAAGGACTT 817

Qy        801 CTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCT 860
              ||| |||||  ||||| || | ||||||  |||||||||||  |   ||||||||||||||
Db        818 CTCTGAGACTTACGAACGCTTCCACACCGAGAGCCTGCAGGGCCGCAGCAAGCAGGAGCT 877

Qy        861 CATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGCATGGAGGACGAGAACAACCG 920
                |    || |||||||| |||||||| | || ||||    |||||| |||| |   |
Db        878 GGTGCGAGACTACCTGGAGCTGGAGAAGCGGCTGTCGCAGGCGGAGGAGGAGACTAGGAG 937

Qy        921 GCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGA---CGACGCGCGTGTGCGGGAGCTGGA 977
              ||||| || |  |     | |    ||      |   |  |   |  |||  ||||||||
Db        938 GCTGCAGCAGCTGCAGGCGTGCACCGGCCAGCAGTCCTGCCGCCAGGTGGAGGAGCTGGC 997

Qy        978 GCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTGACCGAGAACGAACTG 1035
              |||  |  |||| ||  |||| ||||  | ||| |     ||||||  |  ||
Db        998 TGCCGAGGTCCAGAGGCTCCGGACCGAAAACCAGCGGCTTCGTCAGGAGAACCAGATG 1055


RESULT 9
US-10-108-260A-1496/c
; Sequence 1496, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20040005560A1el full length cDNA
;  FILE REFERENCE: H1-A0106
;  CURRENT APPLICATION NUMBER: US/10/108,260A
;  CURRENT FILING DATE:  2002-03-27
;  NUMBER OF SEQ ID NOS: 5458
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1496
;   LENGTH: 2048
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-108-260A-1496

  Query Match            16.0%;  Score 173.2;  DB 12;  Length 2048;
  Best Local Similarity   60.4%;  Pred. No. 7.2e-35;
  Matches  361;  Conservative    0;  Mismatches  198;  Indels   39;  Gaps    3;

Qy        441 GCTGGGGAAGAAAAAACATAAGAGACGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTA 500
              |||||    ||| |||||      | || || || |||    ||  |||| |||  ||| ||
Db       1650 GCTGGCCCGGAAGAAACACCGTCGGCGGCCATCGAAGCGCAAAAGGCACTGGCGACCCTA 1591

Qy        501 CTACAAGCTGAACTGGGAAGAGAAGAAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTC 560
              |    |||||| |||||  |||||  || ||   || ||||  |||||||   | || ||
Db       1590 CCTGGAGCTGAGCTGGGCTGAGAAACAACAGCGGGATGAGAGGCAGAGCCAGAGGGCCTC 1531

Qy        561 AAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTCGCGCCCTATAACACCACGCA 620
                || |||| |  |||||||||||||| |||||||| || || ||| | |||||||| ||
Db       1530 CCGGGTCCGCGAAGAGATGTTCGCCAAAGGCCAGCCCGTGGCCCCCCACAACACCACCCA 1471

Qy        621 GTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAA 680
              |||||| ||| ||||    |||| |||||||||          |||  |  ||| | |||
Db       1470 GTTCCTGATGAATGACAGGGACCCGGAGGAGCC--------CAACTTGGATGTGCCCCAT 1419

Qy        681 GCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGACTTCATGGAAGAAGGGGGTGA 740
              | |  |  || || |     |                           ||||
Db       1418 GGGATCTCCCACCCAGGTTCCAG---------------------------TGGGGAGAG 1387

Qy        741 GGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTT 800
              ||||  |||| ||| ||||||  |||  | |    | | ||||| | || |  ||||||
Db       1386 TGAGGCCGGGGACAGTGATGGGCGGGGCCGAGCGCACGGTGAGTTCCAGCGGAAGGACTT 1327

Qy        801 CTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCT 860
              ||| |||||  ||||| || | ||||||  |||||||||||  |   ||||||||||||||
Db       1326 CTCTGAGACTTACGAACGCTTCCACACCGAGAGCCTGCAGGGCCGCAGCAAGCAGGAGCT 1267

Qy        861 CATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGCATGGAGGACGAGAACAACCG 920
                |    || |||||||| |||||||| | || ||||    |||||| |||| |   |
Db       1266 GGTGCGAGACTACCTGGAGCTGGAGAAGCGGCTGTCGCAGGCGGAGGAGGAGACTAGGAG 1207

Qy        921 GCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGA---CGACGCGCGTGTGCGGGAGCTGGA 977
              ||||| || |  |     | |    ||      |   |  |   |  |||  ||||||||
Db       1206 GCTGCAGCAGCTGCAGGCGTGCACCGGCCAGCAGTCCTGCCGCCAGGTGGAGGAGCTGGC 1147

Qy        978 GCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTGACCGAGAACGAACTG 1035
              |||  |  |||| ||  |||| ||||  | ||| |     ||||||  |  ||
Db       1146 TGCCGAGGTCCAGAGGCTCCGGACCGAAAACCAGCGGCTTCGTCAGGAGAACCAGATG 1089


RESULT 10
US-10-294-804-3
; Sequence 3, Application US/10294804
; Publication No. US20030133948A1
; GENERAL INFORMATION:
;  APPLICANT: Robertson, Erle S.
;  APPLICANT:  Cotter, Murray A.
;  TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
;  TITLE OF INVENTION:  to Genomic Host DNA
;  FILE REFERENCE: UM-03778
;  CURRENT APPLICATION NUMBER: US/10/294,804
;  CURRENT FILING DATE:  2002-11-14
;  PRIOR APPLICATION NUMBER: US/09/410,399
;  PRIOR FILING DATE: 1999-10-01
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 1926
;   TYPE: DNA
;   ORGANISM: Epstein-Barr virus
US-10-294-804-3
```

Query Match 6.3%; Score 68.4; DB 13; Length 1926;
Best Local Similarity 44.0%; Pred. No. 8.2e-08;
Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;



RESULT 11
US-10-291-230-14/c
; Sequence 14, Application US/10291230
; Publication No. US20030108939A1
; GENERAL INFORMATION:
; APPLICANT: Ruffner, Duane E.
; APPLICANT: Pierce, Michael L.
; APPLICANT: Chen, Zhidong
; TITLE OF INVENTION: Directed Antisense Libraries
; FILE REFERENCE: T6678.US.A
; CURRENT APPLICATION NUMBER: US/10/291,230
; CURRENT FILING DATE: 2002-11-07
; PRIOR APPLICATION NUMBER: US 09/647,344
; PRIOR FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: PCT/US99/06742
; PRIOR FILING DATE: 1999-03-28
; PRIOR APPLICATION NUMBER: US 60/079,792
; PRIOR FILING DATE: 1998-03-28
; PRIOR APPLICATION NUMBER: US 60/107,504
; PRIOR FILING DATE: 1998-11-06
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 8705
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: pShuttle
US-10-291-230-14

Query Match 6.3%; Score 68.4; DB 15; Length 8705;
Best Local Similarity 44.0%; Pred. No. 1.1e-07;
Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;



RESULT 12
US-10-291-249-14/c
; Sequence 14, Application US/10291249
; Publication No. US20030119041A1
; GENERAL INFORMATION:
; APPLICANT: Ruffner, Duane E.
; APPLICANT: Pierce, Michael L.
; APPLICANT: Chen, Zhidong
; TITLE OF INVENTION: Directed Antisense Libraries
; FILE REFERENCE: T6678.US.B
; CURRENT APPLICATION NUMBER: US/10/291,249
; CURRENT FILING DATE: 2002-11-07
; PRIOR APPLICATION NUMBER: US 09/647,344
; PRIOR FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: PCT/US99/06742
; PRIOR FILING DATE: 1999-03-28
; PRIOR APPLICATION NUMBER: US 60/079,792

```
;  PRIOR FILING DATE: 1998-03-28
;  PRIOR APPLICATION NUMBER: US 60/107,504
;  PRIOR FILING DATE: 1998-11-06
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
;   LENGTH: 8705
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: pShuttle
US-10-291-249-14

  Query Match             6.3%;  Score 68.4;  DB 15;  Length 8705;
  Best Local Similarity   44.0%;  Pred. No. 1.1e-07;
  Matches  288;  Conservative    0;  Mismatches  366;  Indels    0;  Gaps    0;

Qy        406 GGGGGCGAAGAGGAGTGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGA 465
Db       7982 GGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 7923

Qy        466 CGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAG 525
Db       7922 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGG 7863

Qy        526 AAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCC 585
Db       7862 CAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAG 7803

Qy        586 AAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAG 645
Db       7802 GGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG 7743

Qy        646 GAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGAC 705
Db       7742 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGG 7683

Qy        706 ACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATG 765
Db       7682 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 7623

Qy        766 GGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCAC 825
Db       7622 GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 7563

Qy        826 ACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAG 885
Db       7562 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAG 7503

Qy        886 AAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTG 945
Db       7502 GAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG 7443

Qy        946 GGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAG 1005
Db       7442 GAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 7383

Qy       1006 AACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCG 1059
Db       7382 GGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAG 7329

RESULT 13
US-10-278-751-1
; Sequence 1, Application US/10278751
; Publication No. US20030185890A1
;   GENERAL INFORMATION:
;        APPLICANT: Zuckermann et al.
;        TITLE OF INVENTION: Compositions and Methods for           Polynucleotide
;                            Delivery
;        NUMBER OF SEQUENCES: 4
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Chiron Corporation
;             STREET: 4560 Horton Street
;             CITY: Emeryville
;             STATE: California
;             COUNTRY: U.S.A.
;             ZIP: 94608-2916
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/278,751
;             FILING DATE: 22-Jan-2003
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Fujita, Sharon M.
;             REGISTRATION NUMBER: 38,459
;             REFERENCE/DOCKET NUMBER: 1218.002
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (510) 923-2706
;             TELEFAX: (510) 655-3542
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 9600 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-278-751-1

  Query Match             6.3%;  Score 68.4;  DB 13;  Length 9600;
  Best Local Similarity   44.0%;  Pred. No. 1.1e-07;
  Matches  288;  Conservative    0;  Mismatches  366;  Indels    0;  Gaps    0;

Qy        406 GGGGGCGAAGAGGAGTGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGA 465
Db        734 GGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 793

Qy        466 CGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAG 525
Db        794 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGG 853

Qy        526 AAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCC 585
Db        854 CAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAG 913

Qy        586 AAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAG 645
Db        914 GGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG 973

Qy        646 GAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGAC 705
Db        974 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGG 1033

Qy        706 ACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATG 765
Db       1034 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 1093

Qy        766 GGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCAC 825
Db       1094 GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 1153

Qy        826 ACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAG 885
Db       1154 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAG 1213

Qy        886 AAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTG 945
Db       1214 GAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG 1273

Qy        946 GGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAG 1005
```

```
Db       1274 GAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 1333

Qy       1006 AACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCG 1059

Db       1334 GGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAG 1387


RESULT 14
US-10-050-898-283/c
; Sequence 283, Application US/10050898
; Publication No. US20030175711A1
; GENERAL INFORMATION:
;  APPLICANT: Renner, Wolfgang A.
;  APPLICANT:  Bachmann, Martin
;  APPLICANT:  Tissot, Alain
;  APPLICANT:  Maurer, Patrick
;  APPLICANT:  Lechner, Franziska
;  APPLICANT:  Sebbel, Peter
;  APPLICANT:  Piossek, Christine
;  APPLICANT:  Ortmann, Rainer
;  APPLICANT:  Luond, Rainer
;  APPLICANT:  Staufenbiel, Matthias
;  APPLICANT:  Frey, Peter
;  TITLE OF INVENTION: Molecular Antigen Array
;  FILE REFERENCE: 1700.0190005
;  CURRENT APPLICATION NUMBER: US/10/050,898
;  CURRENT FILING DATE:  2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/262,379
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: US 60/288,549
;  PRIOR FILING DATE: 2001-05-04
;  PRIOR APPLICATION NUMBER: US 60/326,998
;  PRIOR FILING DATE: 2001-10-05
;  PRIOR APPLICATION NUMBER: US 60/331,045
;  PRIOR FILING DATE: 2001-11-07
;  NUMBER OF SEQ ID NOS: 350
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 283
;   LENGTH: 10233
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: pCep-Xa-Fc construct
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (9820)..(9820)
;   OTHER INFORMATION: n is a, c, g, or t
US-10-050-898-283

  Query Match             6.3%;  Score 68.4;  DB 13;  Length 10233;
  Best Local Similarity   44.0%;  Pred. No. 1.1e-07;
  Matches  288;  Conservative    0;  Mismatches  366;  Indels    0;  Gaps    0;

Qy        406 GGGGGCGAAGAGGAGTGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGA 465

Db       4191 GGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 4132

Qy        466 CGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAG 525

Db       4131 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGG 4072

Qy        526 AAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCC 585

Db       4071 CAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAG 4012

Qy        586 AAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAG 645

Db       4011 GGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG 3952

Qy        646 GAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGAC 705

Db       3951 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGG 3892

Qy        706 ACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATG 765

Db       3891 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 3832

Qy        766 GGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCAC 825

Db       3831 GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 3772

Qy        826 ACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAG 885

Db       3771 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAG 3712

Qy        886 AAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTG 945

Db       3711 GAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG 3652

Qy        946 GGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAG 1005

Db       3651 GAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 3592

Qy       1006 AACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCG 1059

Db       3591 GGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAG 3538


RESULT 15
US-10-050-902-283/c
; Sequence 283, Application US/10050902
; Publication No. US20030175290A1
; GENERAL INFORMATION:
;  APPLICANT: Renner, Wolfgang A.
;  APPLICANT:  Bachmann, Martin
;  APPLICANT:  Tissot, Alain
;  APPLICANT:  Maurer, Patrick
;  APPLICANT:  Lechner, Franziska
;  APPLICANT:  Sebbel, Peter
;  APPLICANT:  Piossek, Christine
;  TITLE OF INVENTION: Molecular Antigen Array
;  FILE REFERENCE: 1700.0190004
;  CURRENT APPLICATION NUMBER: US/10/050,902
;  CURRENT FILING DATE:  2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/262,379
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: US 60/288,549
;  PRIOR FILING DATE: 2001-05-04
;  PRIOR APPLICATION NUMBER: US 60/326,998
;  PRIOR FILING DATE: 2001-10-05
;  PRIOR APPLICATION NUMBER: US 60/331,045
;  PRIOR FILING DATE: 2001-11-07
;  NUMBER OF SEQ ID NOS: 350
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 283
;   LENGTH: 10285
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: pCep-Xa-Fc construct
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (9872)..(9872)
;   OTHER INFORMATION: n is a, c, g, or t
US-10-050-902-283

  Query Match             6.3%;  Score 68.4;  DB 13;  Length 10285;
  Best Local Similarity   44.0%;  Pred. No. 1.1e-07;
  Matches  288;  Conservative    0;  Mismatches  366;  Indels    0;  Gaps    0;

Qy        406 GGGGGCGAAGAGGAGTGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAAGAGA 465

Db       4191 GGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 4132

Qy        466 CGCCCGTCCAAGAAGAAGCGGCATTGGAAACCGTACTACAAGCTGAACTGGGAAGAGAAG 525
```

```
Db       4131 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGG 4072

Qy        526 AAAAAGTTCGACGAGAAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCC 585

Db       4071 CAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAG 4012

Qy        586 AAGGGCCAGCCGGTCGCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAG 645

Db       4011 GGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG 3952

Qy        646 GAGGAGCCGGATCTCAAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGAC 705

Db       3951 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGG 3892

Qy        706 ACCAGCGATGACGACTTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATG 765

Db       3891 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG 3832

Qy        766 GGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCAC 825

Db       3831 GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 3772

Qy        826 ACGGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAG 885

Db       3771 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAG 3712

Qy        886 AAGTGCCTCTCGCGCATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTG 945

Db       3711 GAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG 3652

Qy        946 GGTGGCGACGACGCGCGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAG 1005

Db       3651 GAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAG 3592

Qy       1006 AACCTCCAGCTGCTGACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCG 1059

Db       3591 GGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAG 3538
```

Search completed: February 5, 2004, 13:27:01
Job time : 495 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 5, 2004, 13:32:45 ; Search time 4658 Seconds
(without alignments)
3152.975 Million cell updates/sec

Title: US-09-972-758A-2
Perfect score: 1910
Sequence: 1 MAEPFLSEYQHQPQTSNCTG.....LTENELHRQQERAPLSKFGD 359

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09972758/runat_05022004_095003_1189/app_query.fasta_1.519
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09972758_@CGN_1_1_3508_@runat_05022004_095003_1189 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1910	100.0	2086	9	BC006460	BC006460 Homo sapi
2	1910	100.0	2199	6	BD106410	BD106410 Secreted
3	1910	100.0	3624	9	AB021179	AB021179 Homo sapi
c 4	1910	100.0	35050	9	AC142472	AC142472 Homo sapi
5	1910	100.0	38849	9	AC138150	AC138150 Homo sapi
c 6	1910	100.0	100836	2	AC087298	AC087298 Homo sapi
7	1903	99.6	2178	6	BD160090	BD160090 Primer fo
8	1903	99.6	2178	9	AK023624	AK023624 Homo sapi
c 9	1858.5	97.3	142326	2	AC024047	AC024047 Homo sapi
10	1587.5	83.1	3402	10	AY090614	AY090614 Mus muscu
11	1587.5	83.1	111002	10	AL731805	AL731805 Mouse DNA
c 12	1584.5	83.0	213625	2	AC120950	AC120950 Rattus no
13	1577.5	82.6	1488	10	BC022111	BC022111 Mus muscu
c 14	875	45.8	111803	2	AC136172	AC136172 Rattus no
15	714	37.4	523	6	BD058159	BD058159 Secreted
16	629	32.9	461	6	AX331371	AX331371 Sequence
c 17	557.5	29.2	239464	2	AC103079	AC103079 Rattus no
18	557.5	29.2	267540	2	AC107153	AC107153 Rattus no
19	556.5	29.1	1330	6	AX714539	AX714539 Sequence
20	556.5	29.1	1330	9	AK056946	AK056946 Homo sapi
21	550	28.8	96875	9	AC008105	AC008105 Homo sapi
22	550	28.8	223442	10	AL662804	AL662804 Mouse DNA
23	545.5	28.6	1238	9	BC025970	BC025970 Homo sapi
c 24	545	28.5	2048	9	AK096785	AK096785 Homo sapi
25	378	19.8	1980	3	AK115840	AK115840 Ciona int
26	333	17.4	225116	2	AC112538	AC112538 Rattus no
27	300	15.7	111803	2	AC136172	AC136172 Rattus no
28	245	12.8	342	6	BD059439	BD059439 Secreted
29	217	11.4	1561	3	BT001551	BT001551 Drosophil
30	214	11.2	1536	3	AY051786	AY051786 Drosophil
31	205.5	10.8	183439	3	AC007809	AC007809 Drosophil
32	205.5	10.8	232741	3	AE003705	AE003705 Drosophil
c 33	205.5	10.8	237119	2	AC017740	AC017740 Drosophil
34	170	8.9	9121	4	OCTRICHA	Z19092 O.cuniculus
35	165.5	8.7	9344	4	OATRICH	Z18361 O.aries tri
36	165	8.6	6644	4	RABMHCP	M77812 Rabbit myos
37	165	8.6	218667	2	AC095722	AC095722 Rattus no
38	164.5	8.6	4143	5	GSDYNACT	X62773 Gallus gall
39	162.5	8.5	89924	2	AC135973	AC135973 Homo sapi
40	162	8.5	4152	9	HUMMYH9	M81105 Homo sapien
41	162	8.5	4271	9	BC049849	BC049849 Homo sapi
42	162	8.5	7396	6	AX686226	AX686226 Sequence
43	162	8.5	278652	2	AC073809	AC073809 Mus muscu
c 44	162	8.5	314133	2	AC129914	AC129914 Homo sapi
c 45	160	8.4	205288	2	AC140253	AC140253 Mus muscu

ALIGNMENTS

RESULT 1

```
BC006460
LOCUS       BC006460                2086 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, HMBA-inducible, clone MGC:1880 IMAGE:3535529, mRNA,
            complete cds.
ACCESSION   BC006460
VERSION     BC006460.1  GI:13623668
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2086)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-APR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 8 Row: d Column: 1
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 4062855.
FEATURES              Location/Qualifiers
     source           1. .2086
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="LocusID:10614"
                      /db_xref="taxon:9606"
                      /clone="MGC:1880 IMAGE:3535529"
                      /tissue_type="Lung, small cell carcinoma"
                      /clone_lib="NIH_MGC_7"
                      /lab_host="DH10B-R"
                      /note="Vector: pOTB7"
     CDS              606. .1685
                      /codon_start=1
                      /product="HMBA-inducible"
                      /protein_id="AAH06460.1"
                      /db_xref="GI:13623669"
                      /translation="MAEPFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEED
                      SRWQSRAFPQLGGRPGPEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGD
                      FPPPAEVEPTPEAELLAQPCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKK
                      KRHWKPYYKLTWEEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEP
                      DLKTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHT
                      ESLQNMSKQELIKEYLELEKCLSRMEDENNRLRLESKRLGGDDARVRELELELDRLRA
                      ENLQLLTENELHRQQERAPLSKFGD"
BASE COUNT      510 a    505 c    618 g    453 t
ORIGIN

Alignment Scores:
Pred. No.:              1.27e-92        Length:         2086
Score:                  1910.00         Matches:        359
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0

DB:                     9               Gaps:           0

US-09-972-758A-2 (1-359) x BC006460 (1-2086)

Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        606 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 665

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 725

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        726 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 785

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        786 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 845

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 905

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 GGCGACTTCCCGCCGCCGGCAGAAGTGGAGCCGACGCCCGAGGCCGAGCTGCTCGCCCAG 965

Qy        121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 1025

Qy        141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 1085

Qy        161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 1145

Qy        181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 1205

Qy        201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1206 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGACCTC 1265

Qy        221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1266 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 1325

Qy        241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1326 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 1385

Qy        261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1386 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 1445

Qy        281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArg 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1446 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 1505

Qy        301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1506 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 1565

Qy        321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1566 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1625

Qy        341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1626 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGAC 1682

RESULT 2
BD106410
LOCUS       BD106410                2199 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Secreted proteins and polynucleotides encoding them.
ACCESSION   BD106410
VERSION     BD106410.1  GI:23201228
KEYWORDS    JP 2002503955-A/1.
SOURCE      Chlamydia sp.
  ORGANISM  Chlamydia sp.
            Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE   1  (bases 1 to 2199)
  AUTHORS   Jacobs,K., Mccoy,J.M., Lavallie,E.R., Racie,L.A., Merberg,D.,
            Treacy,M., Spaulding,V. and Agostino,M.J.
  TITLE     Secreted proteins and polynucleotides encoding them
  JOURNAL   Patent: JP 2002503955-A 1 05-FEB-2002;
            GENETICS INSTITUTE INC
COMMENT     PN   JP 2002503955-A/1
            PD   05-FEB-2002
            PF   20-MAR-1998 JP 1998545874
            PR   21-MAR-1997 US    08/822167,19-MAR-1998 US    09/044466 PI
            KENNETH JACOBS,JOHN M MCCOY,EDWARD R LAVALLIE,LISA A RACIE, PI
            DAVID MERBERG,
            PI   MAURICE TREACY,VIKKI SPAULDING,MICHAEL J AGOSTINO PC
            C12N15/12,C07K14/47,A61K38/17
            CC   Strandedness: Double;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers.
FEATURES              Location/Qualifiers
     source           1. .2199
                      /organism="Chlamydia sp."
                      /mol_type="genomic DNA"
                      /db_xref="taxon:35827"
BASE COUNT      552 a    511 c    674 g    462 t
ORIGIN

Alignment Scores:
Pred. No.:              1.34e-92        Length:         2199
Score:                  1910.00         Matches:        359
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-972-758A-2 (1-359) x BD106410 (1-2199)

Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        707 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 766

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        767 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 826

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaProGlnLeuGlyGlyArgProGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        827 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 886

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        887 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 946

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        947 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 1006

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1007 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 1066

Qy        121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1067 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 1126

Qy        141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1127 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 1186

Qy        161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1187 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 1246

Qy        181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1247 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 1306

Qy        201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1307 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 1366

Qy        221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1367 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 1426

Qy        241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1427 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 1486

Qy        261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 1546

Qy        281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArg 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 1606

Qy        301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 1666

Qy        321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1726

Qy        341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGAC 1783

RESULT 3
AB021179
LOCUS       AB021179                3624 bp    mRNA    linear   PRI 05-DEC-2000
DEFINITION  Homo sapiens mRNA for HEXIM1 protein, complete cds.
ACCESSION   AB021179
VERSION     AB021179.1  GI:4062855
KEYWORDS    HEXIM1; HEXIM1 protein.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Kusuhara,M., Nagasaki,K., Kimura,K., Maass,N., Manabe,T.,
            Ishikawa,S., Aikawa,M., Miyazaki,K. and Yamaguchi,K.
  TITLE     Cloning of hexamethylene-bis-acetamide-inducible transcript,
            HEXIM1, in human vascular smooth muscle cells
  JOURNAL   Biomed. Res. 20, 273-279 (1999)
REFERENCE   2  (bases 1 to 3624)
  AUTHORS   Kusuhara,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-DEC-1998) Masatoshi Kusuhara, National Defense
            Medical College, The First Department of Internal Medicine; 3-2
            Namiki, Tokorozawa, Saitama 359-8513, Japan
            (E-mail:mkusu@me.ndmc.ac.jp, Tel:+81-42-995-1597,
            Fax:+81-42-996-5200)
FEATURES              Location/Qualifiers
```

```
     source          1. .3624
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_type="coronary artery smooth muscle cell"
     gene            1. .3624
                     /gene="HEXIM1"
     CDS             690. .1769
                     /gene="HEXIM1"
                     /codon_start=1
                     /product="HEXIM1 protein"
                     /protein_id="BAA36166.1"
                     /db_xref="GI:4062856"
                     /translation="MAEPFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEED
                     SRWQSRAFPQLGGRPGPEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGD
                     FPPPAEVEPTPEAELLAQPCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKK
                     KRHWKPYYKLTWEEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEP
                     DLKTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHT
                     ESLQNMSKQELIKEYLELEKCLSRMEDENNRLRLESKRLGGDDARVRELELELDRLRA
                     ENLQLLTENELHRQQERAPLSKFGD"
BASE COUNT      978 a    775 c    964 g    907 t
ORIGIN

Alignment Scores:
Pred. No.:              2.22e-92        Length:         3624
Score:                  1910.00         Matches:        359
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-972-758A-2 (1-359) x AB021179 (1-3624)

Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        690 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 749

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        750 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 809

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        810 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 869

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        870 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 929

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        930 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 989

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        990 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 1049

Qy        121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1050 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 1109

Qy        141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1110 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 1169

Qy        161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1170 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 1229

Qy        181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1230 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 1289

Qy        201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1290 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 1349

Qy        221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1350 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 1409

Qy        241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1410 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 1469

Qy        261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1470 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 1529

Qy        281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArg 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1530 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 1589

Qy        301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1590 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 1649

Qy        321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1650 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1709

Qy        341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1710 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGAC 1766

RESULT 4
AC142472/c
LOCUS       AC142472               35050 bp    DNA     linear   PRI 04-APR-2003
DEFINITION  Homo sapiens chromosome 17, clone XXfos-8119G3, complete sequence.
ACCESSION   AC142472
VERSION     AC142472.1  GI:29540490
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 35050)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone XXfos-8119G3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 35050)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-APR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
```

```
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L29477
                Center clone name: 8119_G_3
            --------------

            Only the last 35.1 kilobases of this clone are being submitted.
            The remainder overlaps accession number AC008105 [WICGR project
            L890].
FEATURES             Location/Qualifiers
     source          1. .35050
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="XXfos-8119G3"
                     /clone_lib="Human Female Fosmid Library"
     repeat_region   14. .320
                     /rpt_family="AluSp"
     repeat_region   1947. .1983
                     /rpt_family="AT_rich"
     repeat_region   3469. .3493
                     /rpt_family="(A)n"
     repeat_region   3993. .4099
                     /rpt_family="C-rich"
     repeat_region   5028. .5069
                     /rpt_family="G-rich"
     repeat_region   complement(5605. .5818)
                     /rpt_family="MIR"
     repeat_region   complement(6064. .6196)
                     /rpt_family="L2"
     repeat_region   complement(6713. .6862)
                     /rpt_family="MIR"
     repeat_region   6863. .7166
                     /rpt_family="AluSc"
     repeat_region   complement(7167. .7225)
                     /rpt_family="MIR"
     repeat_region   8048. .8104
                     /rpt_family="(CCCCG)n"
     repeat_region   8113. .8149
                     /rpt_family="GC_rich"
     repeat_region   9472. .9732
                     /rpt_family="L1MC4"
     repeat_region   9733. .10036
                     /rpt_family="AluSg1"
     repeat_region   10037. .10091
                     /rpt_family="L1MC4"
     repeat_region   complement(10169. .10477)
                     /rpt_family="AluSx"
     repeat_region   10480. .10739
                     /rpt_family="L1MC4"
     repeat_region   10750. .10906
                     /rpt_family="L1MC5"
     repeat_region   complement(10907. .11182)
                     /rpt_family="AluJo"
     repeat_region   11183. .11552
                     /rpt_family="L1MC5"
     repeat_region   complement(11556. .11857)
                     /rpt_family="AluSx"
     repeat_region   complement(11913. .12044)
                     /rpt_family="FLAM_C"
     repeat_region   12108. .12240
                     /rpt_family="FLAM_C"
     repeat_region   12296. .12374
                     /rpt_family="AluS"
     unsure          12353. .12373
                     /note="<30 qual SNGL region"
     unsure          12374
                     /note="unresolved sequence region:
                     100 bp of sequence missing here"
     repeat_region   12375. .12673
                     /rpt_family="AluSc"
     repeat_region   13059. .13232
                     /rpt_family="MIR"
     repeat_region   complement(14041. .14336)
                     /rpt_family="AluSx"
     repeat_region   complement(14518. .14598)
                     /rpt_family="MIR"
     repeat_region   17580. .17884
                     /rpt_family="AluSx"
     repeat_region   17893. .18028
                     /rpt_family="L1MA3"
     repeat_region   complement(18032. .18350)
                     /rpt_family="AluY"
     repeat_region   18360. .18383
                     /rpt_family="AT_rich"
     repeat_region   complement(18387. .18515)
                     /rpt_family="FLAM_C"
     repeat_region   18533. .18707
                     /rpt_family="FRAM"
     repeat_region   18708. .18877
                     /rpt_family="L1MA4"
     repeat_region   19015. .19324
                     /rpt_family="AluYa5"
     repeat_region   20221. .20288
                     /rpt_family="(CCCCG)n"
     repeat_region   20403. .20429
                     /rpt_family="GC_rich"
     repeat_region   complement(21264. .21365)
                     /rpt_family="MIR"
     repeat_region   complement(21662. .21734)
                     /rpt_family="L2"
     repeat_region   complement(22401. .22550)
                     /rpt_family="L2"
     repeat_region   complement(23139. .23338)
                     /rpt_family="MIR"
     repeat_region   complement(23346. .23673)
                     /rpt_family="AluJo"
     repeat_region   complement(23704. .23903)
                     /rpt_family="MIR"
     repeat_region   complement(23996. .24125)
                     /rpt_family="MER53"
     repeat_region   24057. .24138
                     /rpt_family="MER53"
     repeat_region   25014. .25095
                     /rpt_family="(CA)n"
     repeat_region   25274. .25344
                     /rpt_family="MIR3"
     repeat_region   complement(25687. .25742)
                     /rpt_family="L2"
     repeat_region   complement(25847. .25917)
                     /rpt_family="L1MB8"
     repeat_region   25918. .26185
                     /rpt_family="AluJo"
     repeat_region   complement(26186. .26209)
                     /rpt_family="L1MB8"
     repeat_region   26243. .26497
                     /rpt_family="AluYc"
     repeat_region   26500. .26799
                     /rpt_family="AluY"
     repeat_region   complement(26800. .27209)
                     /rpt_family="L1MB4"
     repeat_region   complement(27220. .27528)
                     /rpt_family="AluSx"
     repeat_region   complement(27533. .27592)
                     /rpt_family="AluSq/x"
     repeat_region   complement(27597. .27800)
                     /rpt_family="AluSx"
     repeat_region   complement(27805. .27916)
                     /rpt_family="AluSq/x"
```

```
     repeat_region   complement(27917. .27963)
                     /rpt_family="L1MB4"
     repeat_region   29143. .29261
                     /rpt_family="MER5B"
     repeat_region   complement(29547. .29628)
                     /rpt_family="L2"
     repeat_region   complement(30172. .30304)
                     /rpt_family="MIR"
     repeat_region   30327. .30450
                     /rpt_family="MIR"
     repeat_region   complement(30485. .30611)
                     /rpt_family="AluSg/x"
     repeat_region   30675. .30825
                     /rpt_family="L3b"
     repeat_region   complement(30871. .31136)

Alignment Scores:
Pred. No.:              2.16e-91      Length:         35050
Score:                  1910.00       Matches:        359
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     9             Gaps:           0

US-09-972-758A-2 (1-359) x AC142472 (1-35050)

Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3397 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 3338

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3337 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 3278

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaProGlnLeuGlyGlyArgProGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3277 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 3218

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3217 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 3158

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3157 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 3098

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3097 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 3038

Qy        121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3037 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 2978

Qy        141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2977 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 2918

Qy        161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2917 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 2858

Qy        181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2857 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 2798

Qy        201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2797 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 2738

Qy        221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2737 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 2678

Qy        241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2677 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 2618

Qy        261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2617 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 2558

Qy        281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArg 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2557 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 2498

Qy        301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2497 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 2438

Qy        321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2437 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 2378

Qy        341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2377 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGAC 2321

RESULT 5
AC138150
LOCUS       AC138150               38849 bp    DNA     linear   PRI 14-FEB-2003
DEFINITION  Homo sapiens chromosome 17, clone RP13-890H12, complete sequence.
ACCESSION   AC138150
VERSION     AC138150.4  GI:28376773
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 38849)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP13-890H12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 38849)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-DEC-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 38849)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
```

Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (31-JAN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 4 (bases 1 to 38849)

AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (14-FEB-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Feb 14, 2003 this sequence version replaced gi:28173166.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L28731
Center clone name: 890_H_12

FEATURES Location/Qualifiers

source 1. .38849
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP13-890H12"
/clone_lib="RPCI-13 Human Female BAC"

repeat_region 84. .385
/rpt_family="AluSx"

repeat_region complement(389. .758)
/rpt_family="L1MC5"

repeat_region 759. .1034
/rpt_family="AluJo"

repeat_region complement(1035. .1191)
/rpt_family="L1MC5"

repeat_region complement(1202. .1461)
/rpt_family="L1MC4"

repeat_region 1464. .1772
/rpt_family="AluSx"

repeat_region complement(1850. .1904)
/rpt_family="L1MC4"

repeat_region complement(1905. .2208)
/rpt_family="AluSg1"

repeat_region complement(2209. .2469)
/rpt_family="L1MC4"

repeat_region 3792. .3828
/rpt_family="GC_rich"

repeat_region 3837. .3893
/rpt_family="(CGGGG)n"

repeat_region 4716. .4774
/rpt_family="MIR"

repeat_region complement(4775. .5078)
/rpt_family="AluSc"

repeat_region 5079. .5204
/rpt_family="MIR"

repeat_region 5745. .5877
/rpt_family="L2"

repeat_region 6123. .6336
/rpt_family="MIR3"

repeat_region 6872. .6913
/rpt_family="C-rich"

repeat_region 7842. .7948
/rpt_family="G-rich"

repeat_region 8448. .8472
/rpt_family="(T)n"

repeat_region 9962. .9998
/rpt_family="AT_rich"

repeat_region complement(11625. .11932)
/rpt_family="AluSp"

repeat_region 12103. .12389
/rpt_family="AluSp"

repeat_region complement(12567. .12862)
/rpt_family="AluSx"

repeat_region complement(12863. .13169)
/rpt_family="AluJb"

repeat_region complement(13209. .13491)
/rpt_family="AluY"

repeat_region 13566. .13589
/rpt_family="AT_rich"

repeat_region 13595. .13903
/rpt_family="AluSp"

repeat_region 13904. .14038
/rpt_family="AluSx"

repeat_region 14039. .14352
/rpt_family="AluY"

repeat_region 14353. .14522
/rpt_family="AluSx"

repeat_region 14558. .14685
/rpt_family="FLAM_C"

repeat_region complement(14704. .14854)
/rpt_family="FRAM"

repeat_region 15205. .15307
/rpt_family="AluJb"

repeat_region 15359. .15633
/rpt_family="AluSx"

repeat_region 16074. .16377
/rpt_family="AluSc"

repeat_region 16380. .16558
/rpt_family="AluSg/x"

repeat_region 16743. .16877
/rpt_family="AluSx"

repeat_region 16878. .17181
/rpt_family="AluSp"

repeat_region 17182. .17348
/rpt_family="AluSx"

repeat_region complement(18100. .18399)
/rpt_family="AluSc"

repeat_region complement(18948. .19238)
/rpt_family="AluSx"

repeat_region 22400. .22707
/rpt_family="AluSx"

repeat_region complement(22708. .22845)
/rpt_family="MIR"

repeat_region 22960. .23228

```
                  /rpt_family="AluSx"
     repeat_region   23231. .23524
                  /rpt_family="AluSq"
     repeat_region   23532. .23838
                  /rpt_family="AluSx"
     repeat_region   23843. .24006
                  /rpt_family="AluSg/x"
     repeat_region   25108. .25151

Alignment Scores:
Pred. No.:              2.4e-91       Length:         38849
Score:                  1910.00       Matches:        359
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     9             Gaps:           0

US-09-972-758A-2 (1-359) x AC138150 (1-38849)

Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8544 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 8603

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8604 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 8663

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8664 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 8723

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8724 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 8783

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8784 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 8843

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8844 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 8903

Qy        121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8904 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 8963

Qy        141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8964 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 9023

Qy        161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9024 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 9083

Qy        181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9084 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 9143

Qy        201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9144 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 9203

Qy        221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9204 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 9263

Qy        241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9264 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 9323

Qy        261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       9324 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 9383

Qy        281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArg 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9384 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 9443

Qy        301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9444 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 9503

Qy        321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9504 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 9563

Qy        341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9564 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGAC 9620

RESULT 6
AC087298/c
LOCUS       AC087298               100836 bp    DNA     linear   HTG 20-APR-2002
DEFINITION  Homo sapiens chromosome 17 clone RP11-403G3 map 17, WORKING DRAFT
            SEQUENCE, 15 ordered pieces.
ACCESSION   AC087298
VERSION     AC087298.8  GI:20219131
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 100836)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-403G3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 100836)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-DEC-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 100836)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
```

```
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 20, 2002 this sequence version replaced gi:20177719.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12020
                Center clone name: 403_G_3
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 96277 bases at least Q40
                Consensus quality: 98051 bases at least Q30
                Consensus quality: 98838 bases at least Q20
                Insert size: 140000; agarose-fp
                Insert size: 99436; sum-of-contigs
                Quality coverage: 19.1 in Q20 bases; agarose-fp
                Quality coverage: 26.9 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1     8098: contig of 8098 bp in length
            *     8099     8198: gap of 100 bp
            *     8199     8271: contig of 73 bp in length
            *     8272     8371: gap of 100 bp
            *     8372     9497: contig of 1126 bp in length
            *     9498     9597: gap of 100 bp
            *     9598    10863: contig of 1266 bp in length
            *    10864    10963: gap of 100 bp
            *    10964    12232: contig of 1269 bp in length
            *    12233    12332: gap of 100 bp
            *    12333    13844: contig of 1512 bp in length
            *    13845    13944: gap of 100 bp
            *    13945    15215: contig of 1271 bp in length
            *    15216    15315: gap of 100 bp
            *    15316    18234: contig of 2919 bp in length
            *    18235    18334: gap of 100 bp
            *    18335    21787: contig of 3453 bp in length
            *    21788    21887: gap of 100 bp
            *    21888    25485: contig of 3598 bp in length
            *    25486    25585: gap of 100 bp
            *    25586    40531: contig of 14946 bp in length
            *    40532    40631: gap of 100 bp
            *    40632    52119: contig of 11488 bp in length
            *    52120    52219: gap of 100 bp
            *    52220    64941: contig of 12722 bp in length
            *    64942    65041: gap of 100 bp
            *    65042    79587: contig of 14546 bp in length
            *    79588    79687: gap of 100 bp
            *    79688   100836: contig of 21149 bp in length
FEATURES             Location/Qualifiers
     source          1. .100836
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-403G3"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .8098
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    8199. .8271
                     /note="assembly_fragment"
     misc_feature    8372. .9497
                     /note="assembly_fragment"
     misc_feature    9598. .10863
                     /note="assembly_fragment"
     misc_feature    10964. .12232
                     /note="assembly_fragment"
     misc_feature    12333. .13844
                     /note="assembly_fragment"
     misc_feature    13945. .15215
                     /note="assembly_fragment"
     misc_feature    15316. .18234
                     /note="assembly_fragment"
     misc_feature    18335. .21787
                     /note="assembly_fragment"
     misc_feature    21888. .25485
                     /note="assembly_fragment"
     misc_feature    25586. .40531
                     /note="assembly_fragment"
     misc_feature    40632. .52119
                     /note="assembly_fragment"
     misc_feature    52220. .64941
                     /note="assembly_fragment"
     misc_feature    65042. .79587
                     /note="assembly_fragment"
     misc_feature    79688. .100836
                     /note="assembly_fragment"
BASE COUNT    23052 a  27082 c  27053 g  22249 t   1400 others
ORIGIN

Alignment Scores:
Pred. No.:              6.24e-91        Length:         100836
Score:                  1910.00         Matches:        359
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-972-758A-2 (1-359) x AC087298 (1-100836)

Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70785 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 70726

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70725 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 70666

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70665 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 70606

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70605 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 70546

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      70545 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 70486

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70485 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 70426

Qy        121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70425 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 70366

Qy        141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70365 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 70306

Qy        161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70305 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 70246

Qy        181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70245 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 70186

Qy        201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70185 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 70126

Qy        221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70125 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 70066

Qy        241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70065 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 70006

Qy        261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70005 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 69946

Qy        281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArg 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69945 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 69886

Qy        301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69885 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 69826

Qy        321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69825 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 69766

Qy        341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69765 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGAC 69709
```

```
RESULT 7
BD160090
LOCUS       BD160090                2178 bp    DNA     linear   PAT 17-JAN-2003
DEFINITION  Primer for synthesizing full-length cDNA and use thereof.
ACCESSION   BD160090
VERSION     BD160090.1  GI:27865848
KEYWORDS    JP 2002191363-A/14933.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2178)
  AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
            Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
  TITLE     Primer for synthesizing full-length cDNA and use thereof
  JOURNAL   Patent: JP 2002191363-A 14933 09-JUL-2002;
            HELIX RESEARCH INSTITUTE
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002191363-A/14933
            PD   09-JUL-2002
            PF   28-JUL-2000 JP 2000280990
            PI   TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
            PI   SAITO,
            PI   JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
            PI   KEIICHI NAGAI,TETSUJI OTSUKI
            PC
            C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
             10,
            PC   C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
             Primer for synthesizing full-length cDNA and use thereof FH   Key
                      Location/Qualifiers
            FT   CDS             (716). .(1792).
FEATURES             Location/Qualifiers
     source          1. .2178
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      525 a    513 c    678 g    462 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:                3.12e-92        Length:          2178
Score:                    1903.00         Matches:         358
Percent Similarity:       99.72%          Conservative:    0
Best Local Similarity:    99.72%          Mismatches:      1
Query Match:              99.63%          Indels:          0
DB:                       6               Gaps:            0

US-09-972-758A-2 (1-359) x BD160090 (1-2178)

Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        716 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 775

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        776 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 835

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        836 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 895

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        896 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 955

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        956 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 1015

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1016 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 1075

Qy        121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1076 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 1135

Qy        141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1136 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 1195

Qy        161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1196 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 1255

Qy        181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1256 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 1315

Qy        201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1316 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 1375

Qy        221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240

Db       1376 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 1435

Qy        241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260

Db       1436 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 1495

Qy        261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280

Db       1496 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 1555

Qy        281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArg 300

Db       1556 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 1615

Qy        301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320

Db       1616 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 1675

Qy        321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340

Db       1676 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1735

Qy        341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359

Db       1736 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCCTTCCAAGTTTGGAGAC 1792
```

RESULT 8
AK023624

```
LOCUS       AK023624                2178 bp    mRNA    linear   PRI 01-AUG-2002
DEFINITION  Homo sapiens cDNA FLJ13562 fis, clone PLACE1008080, highly similar
            to Homo sapiens mRNA for HEXIM1 protein.
ACCESSION   AK023624
VERSION     AK023624.1  GI:10435606
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
            Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
            Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
            Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
            Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
            Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2178)
  AUTHORS   Isogai,T. and Otsuki,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
            Helix Research Institute (supported by Japan Key Technology Center
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES             Location/Qualifiers
     source          1. .2178
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="PLACE1008080"
                     /tissue_type="placenta"
                     /clone_lib="PLACE1"
                     /note="cloning vector: pME18SFL3"
BASE COUNT      525 a    513 c    678 g    462 t
ORIGIN
```

Alignment Scores:

Pred. No.:	3.12e-92	Length:	2178
Score:	1903.00	Matches:	358
Percent Similarity:	99.72%	Conservative:	0
Best Local Similarity:	99.72%	Mismatches:	1
Query Match:	99.63%	Indels:	0
DB:	9	Gaps:	0

US-09-972-758A-2 (1-359) x AK023624 (1-2178)

```
Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20

Db        716 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 775

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40

Db        776 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 835

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60

Db        836 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 895

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80

Db        896 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 955

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100

Db        956 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 1015

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120

Db       1016 GGCGACTTCCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 1075

Qy        121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140

Db       1076 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 1135

Qy        141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160

Db       1136 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 1195

Qy        161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180

Db       1196 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 1255

Qy        181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200

Db       1256 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 1315

Qy        201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220

Db       1316 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 1375

Qy        221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240

Db       1376 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 1435

Qy        241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260

Db       1436 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 1495

Qy        261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280

Db       1496 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 1555

Qy        281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArg 300

Db       1556 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 1615
```

```
Qy        301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1616 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 1675

Qy        321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1676 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 1735

Qy        341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359
              ||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Db       1736 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCCTTCCAAGTTTGGAGAC 1792

RESULT 9
AC024047/c
LOCUS       AC024047              142326 bp    DNA     linear   HTG 07-JUL-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-403G3, WORKING DRAFT
            SEQUENCE, 28 unordered pieces.
ACCESSION   AC024047
VERSION     AC024047.3  GI:8570008
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 142326)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 142326)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 16, 2000 this sequence version replaced gi:7109599.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0403G03
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 127942 bases at least Q40
            Consensus quality: 133308 bases at least Q30
            Consensus quality: 135734 bases at least Q20
            Insert size: 147000; agarose-fp
            Insert size: 139626; sum-of-contigs
            Quality coverage: 4.06 in Q20 bases; agarose-fp
            Quality coverage: 4.33 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 28 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1575: contig of 1575 bp in length
            *     1576     1675: gap of unknown length
            *     1676     3114: contig of 1439 bp in length
            *     3115     3214: gap of unknown length
            *     3215     5389: contig of 2175 bp in length
            *     5390     5489: gap of unknown length
            *     5490     7129: contig of 1640 bp in length
            *     7130     7229: gap of unknown length
            *     7230     9240: contig of 2011 bp in length
            *     9241     9340: gap of unknown length
            *     9341    11177: contig of 1837 bp in length
            *    11178    11277: gap of unknown length
            *    11278    13113: contig of 1836 bp in length
            *    13114    13213: gap of unknown length
            *    13214    16526: contig of 3313 bp in length
            *    16527    16626: gap of unknown length
            *    16627    19864: contig of 3238 bp in length
            *    19865    19964: gap of unknown length
            *    19965    22667: contig of 2703 bp in length
            *    22668    22767: gap of unknown length
            *    22768    26650: contig of 3883 bp in length
            *    26651    26750: gap of unknown length
            *    26751    29303: contig of 2553 bp in length
            *    29304    29403: gap of unknown length
            *    29404    32073: contig of 2670 bp in length
            *    32074    32173: gap of unknown length
            *    32174    37141: contig of 4968 bp in length
            *    37142    37241: gap of unknown length
            *    37242    40738: contig of 3497 bp in length
            *    40739    40838: gap of unknown length
            *    40839    45166: contig of 4328 bp in length
            *    45167    45266: gap of unknown length
            *    45267    49501: contig of 4235 bp in length
            *    49502    49601: gap of unknown length
            *    49602    55647: contig of 6046 bp in length
            *    55648    55747: gap of unknown length
            *    55748    61532: contig of 5785 bp in length
            *    61533    61632: gap of unknown length
            *    61633    68005: contig of 6373 bp in length
            *    68006    68105: gap of unknown length
            *    68106    75211: contig of 7106 bp in length
            *    75212    75311: gap of unknown length
            *    75312    81560: contig of 6249 bp in length
            *    81561    81660: gap of unknown length
            *    81661    88524: contig of 6864 bp in length
            *    88525    88624: gap of unknown length
            *    88625    95519: contig of 6895 bp in length
            *    95520    95619: gap of unknown length
            *    95620   105148: contig of 9529 bp in length
            *   105149   105248: gap of unknown length
            *   105249   115269: contig of 10021 bp in length
            *   115270   115369: gap of unknown length
            *   115370   128646: contig of 13277 bp in length
            *   128647   128746: gap of unknown length
            *   128747   142326: contig of 13580 bp in length.
FEATURES             Location/Qualifiers
     source          1. .142326
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /clone="RP11-403G3"
BASE COUNT    32503 a  37098 c  37256 g  32736 t   2733 others
ORIGIN

Alignment Scores:
Pred. No.:                4.66e-88        Length:          142326
Score:                    1858.50         Matches:         357
Percent Similarity:       99.44%          Conservative:    0
Best Local Similarity:    99.44%          Mismatches:      2
Query Match:              97.30%          Indels:          2
DB:                       2               Gaps:            0

US-09-972-758A-2 (1-359) x AC024047 (1-142326)

Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     108505 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 108446

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     108445 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 108386

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     108385 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 108326

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     108325 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 108266

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     108265 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 108206

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
                 |||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db     108205 GC-GACTTCCCGCG-CCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 108148

Qy        121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     108147 CCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGGCGAAGAGGAG 108088

Qy        141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     108087 TGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATAGGAGACGCCCGTCCAAGAAG 108028

Qy        161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     108027 AAGCGGCATTGGAAACCGTACTACAAGCTGACCTGGGAAGAGAAGAAAAAGTTCGACGAG 107968

Qy        181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     107967 AAACAGAGCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTC 107908

Qy        201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     107907 GCGCCCTATAACACCACGCAGTTCCTCATGGATGATCACGACCAGGAGGAGCCGGATCTC 107848

Qy        221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     107847 AAAACCGGCCTGTACTCCAAGCGGGCCGCCGCCAAATCCGACGACACCAGCGATGACGAC 107788

Qy        241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     107787 TTCATGGAAGAAGGGGGTGAGGAGGATGGGGGCAGCGATGGGATGGGAGGGGACGGCAGC 107728

Qy        261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     107727 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACACGGAGAGCCTGCAG 107668

Qy        281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArg 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     107667 AACATGAGCAAGCAGGAGCTCATCAAGGAGTACCTGGAACTGGAGAAGTGCCTCTCGCGC 107608

Qy        301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     107607 ATGGAGGACGAGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACGCG 107548

Qy        321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     107547 CGTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGAGAACCTCCAGCTGCTG 107488

Qy        341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     107487 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCGCTTTCCAAGTTTGGAGAC 107431

RESULT 10
AY090614
LOCUS       AY090614               3402 bp    mRNA    linear   ROD 18-JUL-2002
DEFINITION  Mus musculus cardiac lineage protein 1 (Clp1) mRNA, complete cds.
ACCESSION   AY090614
VERSION     AY090614.1  GI:20135640
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3402)
  AUTHORS   Huang,F., Wagner,M. and Siddiqui,M.A.Q.
  TITLE     Structure, expression, and functional characterization of the mouse
            CLP-1 gene
  JOURNAL   Gene 292 (1-2), 245-259 (2002)
  MEDLINE   22114986
   PUBMED   12119119
REFERENCE   2  (bases 1 to 3402)
  AUTHORS   Huang,F., Wagner,M. and Siddiqui,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2002) Anatomy and Cell Biology, State University
            of New York Health Science Center, 450 Clarkson Avenue, Brooklyn,
            NY 11203, USA
FEATURES             Location/Qualifiers
     source          1. .3402
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="129/SvJ"
                     /db_xref="taxon:10090"
     gene            1. .3402
                     /gene="Clp1"
     CDS             596. .1666
                     /gene="Clp1"
                     /note="nuclear protein; CLP-1"
                     /codon_start=1
                     /product="cardiac lineage protein 1"
                     /protein_id="AAM09026.1"
                     /db_xref="GI:20135641"
                     /translation="MAEPLLTEHQHQPQTSNCTGAAVVHEEHTSERPPSAEERVPKED
                     SRWQSRASLQSGSRPGQEGEGGLKHQLPPLQTNACPELSSLEKGEKGQNGEDLSTGGA
                     SPSAEGEPMSESLVQPGHDSEATKQEAPAAGGEEPWGQQQRQLGKKKHRRRPSKKKRH
                     WKPYYKLTWEEKKKFDEKQSLRASRVRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLK
                     TGLYPKRAAAKSDDTSDEDFVEEAGEEDGGSDGMGGDGSEFLQRDFSETYERYHAESL
                     QNMSKQELIKEYLELEKCLSRKEDENNRLRLESKRLGGVDARVRELELELDRLRAENL
                     QLLTENELHRQQERAPLSKFGD"
BASE COUNT      920 a    771 c    890 g    821 t
ORIGIN

Alignment Scores:
Pred. No.:                2.32e-75        Length:          3402
Score:                    1587.50         Matches:         308
Percent Similarity:       88.86%          Conservative:    11
Best Local Similarity:    85.79%          Mismatches:      37
Query Match:              83.12%          Indels:          3
DB:                       10              Gaps:            2

US-09-972-758A-2 (1-359) x AY090614 (1-3402)

Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
              ||||||||||||   |||:::|||:::|||||||||||||||||||||||||||||||||
Db        596 ATGGCCGAGCCACTCTTGACAGAACATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 655

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              ||||||   |||   ||||||          ||||||||||||   ||||||||||||||||
Db        656 GCTGCTGTTGTCCATGAAGAGCATACCTCTGAGCGCCCCCCAAGCGCGGAGGAACGGGTG 715

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
              |||:::|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        716 CCCAAGGAGGACAGTAGGTGGCAATCGAGAGCGTCCTTGCAGTCCGGTAGCCGTCCAGGG 775

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
              ||||||||||||   |||:::   |||   ||||||||||||||||   |||||||||
Db        776 CAGGAGGGAGAAGGGGGCCTGAAGCACCAGCTGCCCCCATTGCAGACGAATGCCTGTCCA 835

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
              |||   |||   |||   :::|||||||||||||||||||:::|||   |||   |||
Db        836 GAATTGAGCTCCCTGGAGAAGGGCGAGAAGGGCCAGAATGGGGAGGACTTATCCACTGGC 895

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
```

```
Db        896 GGT---GCCTCCCCGTCGGCGGAGGGAGAGCCGATGTCAGAGTCC------CTCGTGCAG 946

Qy        121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140

Db        947 CCAGGTCATGACTCGGAGGCCACCAAGCAGGAGGCTCCTGCCGCCGGAGGCGAGGAGCCA 1006

Qy        141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160

Db       1007 TGGGGACAGCAACAGAGACAGCTGGGCAAGAAAAAACATCGGAGACGCCCCTCAAAGAAG 1066

Qy        161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180

Db       1067 AAGCGGCATTGGAAGCCCTACTACAAGCTGACTTGGGAGGAGAAGAAAAAGTTCGACGAG 1126

Qy        181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200

Db       1127 AAGCAGAGCCTGCGAGCTTCGCGGGTTCGAGCCGAGATGTTCGCCAAGGGCCAGCCAGTT 1186

Qy        201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220

Db       1187 GCGCCCTATAACACCACGCAGTTCCTCATGGATGACCACGATCAGGAGGAGCCTGATCTC 1246

Qy        221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240

Db       1247 AAAACCGGCCTTTACCCCAAGCGGGCAGCCGCCAAATCCGACGACACCAGCGATGAGGAT 1306

Qy        241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260

Db       1307 TTTGTGGAAGAAGCTGGTGAGGAAGACGGAGGCAGCGATGGCATGGGAGGGGACGGCAGC 1366

Qy        261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280

Db       1367 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACGCCGAGAGCCTGCAG 1426

Qy        281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArg 300

Db       1427 AACATGAGCAAGCAGGAGCTCATCAAAGAGTACCTGGAGCTGGAGAAGTGCCTCTCCCGC 1486

Qy        301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320

Db       1487 AAGGAGGACGAAAATAACCGGCTGCGGCTGGAAAGCAAGCGGCTGGGAGGCGTCGACGCG 1546

Qy        321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340

Db       1547 CGAGTGCGGGAGCTCGAGCTAGAGCTGGACCGGCTGCGCGCTGAGAACCTCCAGCTGCTG 1606

Qy        341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359

Db       1607 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCTCTTTCCAAGTTCGGCGAC 1663
```

RESULT 11
AL731805
LOCUS AL731805 111002 bp DNA linear ROD 21-JUN-2002
DEFINITION Mouse DNA sequence from clone RP23-358E19 on chromosome 11, complete sequence.
ACCESSION AL731805
VERSION AL731805.8 GI:21540125
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 111002)
AUTHORS Dunn,M.
TITLE Direct Submission
JOURNAL Submitted (21-JUN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Jun 23, 2002 this sequence version replaced gi:21531483.
---------- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-358E19 is from the RPCI-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.
FEATURES Location/Qualifiers
source 1. .111002
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-358E19"
/clone_lib="RPCI-23"
BASE COUNT 28051 a 27579 c 27135 g 28237 t
ORIGIN

Alignment Scores:

Pred. No.:	7.68e-74	Length:	111002
Score:	1587.50	Matches:	308
Percent Similarity:	88.86%	Conservative:	11
Best Local Similarity:	85.79%	Mismatches:	37
Query Match:	83.12%	Indels:	3
DB:	10	Gaps:	2

US-09-972-758A-2 (1-359) x AL731805 (1-111002)

```
Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20

Db      91440 ATGGCCGAGCCACTCTTGACAGAACATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 91499

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40

Db      91500 GCTGCTGTTGTCCATGAAGAGCATACCTCTGAGCGCCCCCCAAGCGCGGAGGAACGGGTG 91559

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60

Db      91560 CCCAAGGAGGACAGTAGGTGGCAATCGAGAGCGTCCTTGCAGTCCGGTAGCCGTCCAGGG 91619

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80

Db      91620 CAGGAGGGAGAAGGGGGCCTGAAGCACCAGCTGCCCCCATTGCAGACGAATGCCTGTCCA 91679

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100

Db      91680 GAATTGAGCTCCCTGGAGAAGGGCGAGAAGGGCCAGAATGGGGAGGACTTATCCACTGGC 91739

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120

Db      91740 GGT---GCCTCCCCGTCGGCGGAGGGAGAGCCGATGTCAGAGTCC------CTCGTGCAG 91790

Qy        121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140

Db      91791 CCAGGTCATGACTCGGAGGCCACCAAGCAGGAGGCTCCTGCCGCCGGAGGCGAGGAGCCA 91850
```

```
Qy        141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91851 TGGGGACAGCAACAGAGACAGCTGGGCAAGAAAAAACATCGGAGACGCCCCTCAAAGAAG 91910

Qy        161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91911 AAGCGGCATTGGAAGCCCTACTACAAGCTGACTTGGGAGGAGAAGAAAAAGTTCGACGAG 91970

Qy        181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
              ||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db      91971 AAGCAGAGCCTGCGAGCTTCGCGGGTTCGAGCCGAGATGTTCGCCAAGGGCCAGCCAGTT 92030

Qy        201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      92031 GCGCCCTATAACACCACGCAGTTCCTCATGGATGACCACGATCAGGAGGAGCCTGATCTC 92090

Qy        221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240
              ||||||||||||||   ||||||||||||||||||||||||||||||:::|||
Db      92091 AAAACCGGCCTTTACCCCAAGCGGGCAGCCGCCAAATCCGACGACACCAGCGATGAGGAT 92150

Qy        241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260
              |||:::||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db      92151 TTTGTGGAAGAAGCTGGTGAGGAAGACGGAGGCAGCGATGGCATGGGAGGGGACGGCAGC 92210

Qy        261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280
              ||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db      92211 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACGCCGAGAGCCTGCAG 92270

Qy        281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArg 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      92271 AACATGAGCAAGCAGGAGCTCATCAAAGAGTACCTGGAGCTGGAGAAGTGCCTCTCCCGC 92330

Qy        301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320
              ||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db      92331 AAGGAGGACGAAAATAACCGGCTGCGGCTGGAAAGCAAGCGGCTGGGAGGCGTCGACGCG 92390

Qy        321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      92391 CGAGTGCGGGAGCTCGAGCTAGAGCTGGACCGGCTGCGCGCTGAGAACCTCCAGCTGCTG 92450

Qy        341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      92451 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCTCTTTCCAAGTTCGGCGAC 92507
```

RESULT 12
AC120950/c
LOCUS AC120950 213625 bp DNA linear HTG 19-NOV-2002
DEFINITION Rattus norvegicus clone CH230-456H21, WORKING DRAFT SEQUENCE, 5 unordered pieces.
ACCESSION AC120950
VERSION AC120950.4 GI:25091749
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 213625)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 213625)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (14-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 213625)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 19, 2002 this sequence version replaced gi:23907976.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GWRQ
Center clone name: CH230-456H21
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 196084 bases at least Q40

```
            Consensus quality: 197843 bases at least Q30
            Consensus quality: 198926 bases at least Q20
            Estimated insert size: 201713; sum-of-contigs estimation
            Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
          --------------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 5 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1    58094: contig of 58094 bp in length
          *   58095    58194: gap of unknown length
          *   58195   139497: contig of 81303 bp in length
          *  139498   139597: gap of unknown length
          *  139598   208746: contig of 69149 bp in length
          *  208747   208846: gap of unknown length
          *  208847   210051: contig of 1205 bp in length
          *  210052   210151: gap of unknown length
          *  210152   213625: contig of 3474 bp in length.
FEATURES             Location/Qualifiers
     source          1. .213625
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-456H21"
     misc_feature    1. .1190
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    47586. .47846
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end_sequence:BZ240544"
BASE COUNT    51942 a  49325 c  49186 g  48940 t  14232 others
ORIGIN

Alignment Scores:
Pred. No.:              2.14e-73        Length:         213625
Score:                  1584.50         Matches:        307
Percent Similarity:     88.58%          Conservative:   11
Best Local Similarity:  85.52%          Mismatches:     38
Query Match:            82.96%          Indels:         3
DB:                     2               Gaps:           2

US-09-972-758A-2 (1-359) x AC120950 (1-213625)

Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
              ||||||||||||   |||||||||:::||||||||||||||||||||||||||||||||
Db     119733 ATGGCTGAGCCACTCTTGTCAGAACATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 119674

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              ||||||   |||   ||||||   |||   ||||||||||||   ||||||||||||||||
Db     119673 GCTGCTGTTGTCCATGAGGAGCAGAACTCTGAGCGCCCCCCAAGCGCGGAGGAACGGGTG 119614

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaProGlnLeuGlyGlyArgProGly 60
              |||:::|||||||||||||||||||||||||||      |||   |||   |||||||||
Db     119613 CCTAAGGAGGACAGTAGGTGGCAATCGAGAGCGTCCTTGCAGTCGGGTAGCCGTCCAGGG 119554

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
              ||||||:::|||   |||:::   |||   ||||||||||||||   |||||||||
Db     119553 CAGGAGGGAGACGGGGGCCTGAAGCACCAGCTGCCCCCATTGCAGACGAATGCCTGTCCA 119494

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
              |||   |||||||||   :::|||||||||||||||||||:::|||   |||   |||
Db     119493 GAATTGAGCTGCTTGGAAAAGGGCGAGAAGGGCCAGAATGGGGAGGACTTATCCACTGGC 119434

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
              |||         |||   ||||||   ||||||      |||:::      |||   |||

Db     119433 GGT---GCCTCGCCTTCGGCGGAGGGGGAGCCGATGTCAGAGTCC------CTCGTGCAA 119383

Qy         121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140
               |||   ||||||||||||||:::||||||   ||||||   ||||||   ||||||
Db     119382 CCAGGTCATGACTCGGAGGCTACTAAGCTGGAGGCTCCTGTCGCTGGAGTAGAGGAGCCA 119323

Qy         141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     119322 TGGGGACAGCAACAGAGACAGCTGGGCAAGAAAAAACACCGGAGACGCCCCTCCAAGAAG 119263

Qy         161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     119262 AAGCGACATTGGAAGCCCTATTACAAGCTGACTTGGGAGGAGAAGAAAAAGTTCGACGAG 119203

Qy         181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
               ||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db     119202 AAGCAAAGCCTGCGAGCTTCGCGAGTTCGAGCCGAGATGTTCGCCAAGGGCCAGCCGGTT 119143

Qy         201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     119142 GCGCCCTATAACACCACGCAGTTCCTCATGGATGACCATGATCAGGAGGAGCCTGATCTC 119083

Qy         221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240
               ||||||||||||||   ||||||||||||||||||||||||||||||||||:::|||
Db     119082 AAAACCGGCCTTTACCCTAAGCGGGCAGCCGCCAAATCCGACGACACCAGCGATGAGGAT 119023

Qy         241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260
               |||:::||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db     119022 TTCGTGGAAGAAGCTGGTGAGGAAGACGGAGGCAGCGATGGCATGGGAGGGGACGGCAGC 118963

Qy         261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280
               ||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db     118962 GAGTTTCTGCAGCGGGACTTTTCCGAGACTTACGAGCGGTACCACGCCGAGAGCCTGCAG 118903

Qy         281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArg 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     118902 AACATGAGCAAGCAGGAGCTCATCAAAGAGTACCTGGAGCTGGAGAAGTGCCTCTCCCGC 118843

Qy         301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320
               ||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db     118842 AAGGAGGACGAAAATAACCGGCTGCGGTTGGAAAGCAAGCGGTTGGGTGGCGTCGATGCG 118783

Qy         321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340
               ||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db     118782 CGAGTGCGGGAGCTGGAGCTAGAGCTGGACCGACTGCGCGCCGAGAACCGCCAGCTGCTG 118723

Qy         341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359
               |||||||||||||||||||||||||||||||||||||   ||||||||||||||
Db     118722 ACCGAAAACGAACTGCACCGGCAGCAGGAGCGAGCGCCTCCTTCCAAGTTCGGCGAC 118666

RESULT 13
BC022111
LOCUS       BC022111                1488 bp    mRNA    linear   ROD 16-APR-2003
DEFINITION  Mus musculus cardiac lineage protein 1, mRNA (cDNA clone MGC:36249
            IMAGE:5067620), complete cds.
ACCESSION   BC022111
VERSION     BC022111.1  GI:18381039
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1488)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
```

Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

MEDLINE 22388257

PUBMED 12477932

REFERENCE 2 (bases 1 to 1488)

AUTHORS Strausberg,R.

TITLE Direct Submission

JOURNAL Submitted (24-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov

COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 57 Row: a Column: 15
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.

FEATURES Location/Qualifiers

source 1. .1488
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="MGC:36249 IMAGE:5067620"
/tissue_type="Mammary tumor metastatized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."
/clone_lib="NCI_CGAP_Lu29"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

gene 1. .1488
/gene="Clp1-pending"
/note="synonyms: HIS1, CLP-1"
/db_xref="LocusID:192231"
/db_xref="MGI:2385923"

CDS 448. .1188
/codon_start=1
/product="Clp1-pending protein"
/protein_id="AAH22111.1"
/db_xref="GI:18381040"
/db_xref="LocusID:192231"
/translation="MSESLVQPGHDSEATKQEAPAAGGEEPWGQQQRQLGKKKHRRRP
SKKKRHWKPYYKLTWEEKKKFDEKQSLRASRVRAEMFAKGQPVAPYNTTQFLMDDHDQ
EEPDLKTGLYPKRAAAKSDDTSDEDFVEEAGEEDGGSDGMGGDGSEFLQRDFSETYER
YHAESLQNMSKQELIKEYLELEKCLSRKEDENNRLRLESKRLGGVDARVRELELELDR
LRAENLQLLTENELHRQQERAPLSKFGD"

BASE COUNT 385 a 371 c 449 g 283 t

ORIGIN

Alignment Scores:

Pred. No.:	3.42e-75	Length:	1488
Score:	1577.50	Matches:	305
Percent Similarity:	88.86%	Conservative:	14
Best Local Similarity:	84.96%	Mismatches:	37
Query Match:	82.59%	Indels:	3
DB:	10	Gaps:	2

US-09-972-758A-2 (1-359) x BC022111 (1-1488)

```
Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
              :::::::::|||   |||:::|||:::||||||||||||||||||||||||||||||||
Db        118 GTCTCACAGCCACTCTTGACAGAACATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 177

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              ||||||   |||  ||||||          ||||||||||||   ||||||||||||||||
Db        178 GCTGCTGTTGTCCATGAAGAGCATACCTCTGAGCGCCCCCCAAGCGCGGAGGAACGGGTG 237

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
              |||:::||||||||||||||||||||||||||||      |||  |||  |||||||||
Db        238 CCCAAGGAGGACAGTAGGTGGCAATCGAGAGCGTCCTTGCAGTCCGGTAGCCGTCCAGGG 297

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
              ||||||||||||   |||:::   |||  ||||||||||||||||   |||||||||
Db        298 CAGGAGGGAGAAGGGGGCCTGAAGCACCAGCTGCCCCCATTGCAGACGAATGCCTGTCCA 357

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
              |||   |||  |||   :::|||||||||||||||||||:::|||   |||   |||
Db        358 GAATTGAGCTCCCTGGAGAAGGGCGAGAAGGGCCAGAATGGGGAGGACTTATCCACTGGC 417

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
              |||          |||  ||||||  ||||||      |||:::      |||  |||
Db        418 GGT---GCCTCCCCGTCGGCGGAGGGAGAGCCGATGTCAGAGTCC------CTCGTGCAG 468

Qy        121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140
              |||   ||||||||||||||:::|||      ||||||||||||||||||||||
Db        469 CCAGGTCATGACTCGGAGGCCACCAAGCAGGAGGCTCCTGCCGCCGGAGGCGAGGAGCCA 528

Qy        141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        529 TGGGGACAGCAACAGAGACAGCTGGGCAAGAAAAAACATCGGAGACGCCCCTCAAAGAAG 588

Qy        161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        589 AAGCGGCATTGGAAGCCCTACTACAAGCTGACTTGGGAGGAGAAGAAAAAGTTCGACGAG 648

Qy        181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
              ||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||
Db        649 AAGCAGAGCCTGCGAGCTTCGCGGGTTCGAGCCGAGATGTTCGCCAAGGGCCAGCCAGTT 708

Qy        201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        709 GCGCCCTATAACACCACGCAGTTCCTCATGGATGACCACGATCAGGAGGAGCCTGATCTC 768

Qy        221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240
              ||||||||||||||   |||||||||||||||||||||||||||||||||:::|||
Db        769 AAAACCGGCCTTTACCCCAAGCGGGCAGCCGCCAAATCCGACGACACCAGCGATGAGGAT 828

Qy        241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260
              |||:::||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db        829 TTTGTGGAAGAAGCTGGTGAGGAAGACGGAGGCAGCGATGGCATGGGAGGGGACGGCAGC 888

Qy        261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280
              ||||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db        889 GAGTTTCTGCAGCGGGACTTCTCGGAGACGTACGAGCGGTACCACGCCGAGAGCCTGCAG 948

Qy        281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArg 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        949 AACATGAGCAAGCAGGAGCTCATCAAAGAGTACCTGGAGCTGGAGAAGTGCCTCTCCCGC 1008
```

```
Qy        301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db       1009 AAGGAGGACGAAAATAACCGGCTGCGGCTGGAAAGCAAGCGGCTGGGAGGCGTCGACGCG 1068

Qy        321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1069 CGAGTGCGGGAGCTCGAGCTAGAGCTGGACCGGCTGCGCGCTGAGAACCTCCAGCTGCTG 1128

Qy        341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1129 ACCGAGAACGAACTGCACCGGCAGCAGGAGCGAGCGCCTCTTTCCAAGTTCGGCGAC 1185

RESULT 14
AC136172/c
LOCUS       AC136172               111803 bp    DNA     linear   HTG 30-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-226J2, *** SEQUENCING IN PROGRESS
            ***, 53 unordered pieces.
ACCESSION   AC136172
VERSION     AC136172.1  GI:24421642
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 111803)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., Mcleod,M., Mcneill,T., Meenen,E., Milosavljevic,A.,
            Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
            Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
            Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G.,
            Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
            Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A.,
            Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
            Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
            Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
            Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
            Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
            Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E.,
            Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
            Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
            Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
            Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
            Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R.,
            Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
            Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
            Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
            Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 111803)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     ---------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
              Center project name: KCVI
              Center clone name: CH230-226J2
            ---------- Summary Statistics
              Sequencing vector: Plasmid;
              Chemistry: Dye-terminator Big Dye: 100% of reads
              Assembly program: Phrap; version 0.990329
              Consensus quality: 67362 bases at least Q40
              Consensus quality: 72560 bases at least Q30
              Consensus quality: 77367 bases at least Q20
            ----------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 53 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1244: contig of 1244 bp in length
            *     1245     1344: gap of unknown length
            *     1345     2653: contig of 1309 bp in length
            *     2654     2753: gap of unknown length
            *     2754     4274: contig of 1521 bp in length
            *     4275     4374: gap of unknown length
            *     4375     5978: contig of 1604 bp in length
            *     5979     6078: gap of unknown length
            *     6079     7135: contig of 1057 bp in length
            *     7136     7235: gap of unknown length
            *     7236     8520: contig of 1285 bp in length
            *     8521     8620: gap of unknown length
            *     8621    10172: contig of 1552 bp in length
            *    10173    10272: gap of unknown length
            *    10273    11379: contig of 1107 bp in length
            *    11380    11479: gap of unknown length
            *    11480    12576: contig of 1097 bp in length
            *    12577    12676: gap of unknown length
            *    12677    13863: contig of 1187 bp in length
            *    13864    13963: gap of unknown length
            *    13964    14987: contig of 1024 bp in length
            *    14988    15087: gap of unknown length
            *    15088    16573: contig of 1486 bp in length
            *    16574    16673: gap of unknown length
            *    16674    17896: contig of 1223 bp in length
            *    17897    17996: gap of unknown length
            *    17997    19475: contig of 1479 bp in length
            *    19476    19575: gap of unknown length
            *    19576    20884: contig of 1309 bp in length
            *    20885    20984: gap of unknown length
            *    20985    22487: contig of 1503 bp in length
            *    22488    22587: gap of unknown length
            *    22588    24747: contig of 2160 bp in length
            *    24748    24847: gap of unknown length
            *    24848    26172: contig of 1325 bp in length
            *    26173    26272: gap of unknown length
            *    26273    27802: contig of 1530 bp in length
            *    27803    27902: gap of unknown length
```

```
*    27903    29918: contig of 2016 bp in length
*    29919    30018: gap of unknown length
*    30019    31165: contig of 1147 bp in length
*    31166    31265: gap of unknown length
*    31266    32793: contig of 1528 bp in length
*    32794    32893: gap of unknown length
*    32894    34401: contig of 1508 bp in length
*    34402    34501: gap of unknown length
*    34502    35644: contig of 1143 bp in length
*    35645    35744: gap of unknown length
*    35745    36968: contig of 1224 bp in length
*    36969    37068: gap of unknown length
*    37069    38349: contig of 1281 bp in length
*    38350    38449: gap of unknown length
*    38450    41409: contig of 2960 bp in length
*    41410    41509: gap of unknown length
*    41510    43066: contig of 1557 bp in length
*    43067    43166: gap of unknown length
*    43167    44827: contig of 1661 bp in length
*    44828    44927: gap of unknown length
*    44928    46592: contig of 1665 bp in length
*    46593    46692: gap of unknown length
*    46693    49255: contig of 2563 bp in length
*    49256    49355: gap of unknown length
*    49356    51247: contig of 1892 bp in length
*    51248    51347: gap of unknown length
*    51348    53114: contig of 1767 bp in length
*    53115    53214: gap of unknown length
*    53215    55332: contig of 2118 bp in length
*    55333    55432: gap of unknown length
*    55433    57427: contig of 1995 bp in length
*    57428    57527: gap of unknown length
*    57528    59606: contig of 2079 bp in length
*    59607    59706: gap of unknown length
*    59707    61138: contig of 1432 bp in length
*    61139    61238: gap of unknown length
*    61239    63710: contig of 2472 bp in length
*    63711    63810: gap of unknown length
*    63811    66371: contig of 2561 bp in length
*    66372    66471: gap of unknown length
*    66472    70029: contig of 3558 bp in length
*    70030    70129: gap of unknown length
*    70130    72934: contig of 2805 bp in length
*    72935    73034: gap of unknown length
*    73035    76024: contig of 2990 bp in length
*    76025    76124: gap of unknown length
*    76125    79346: contig of 3222 bp in length
*    79347    79446: gap of unknown length
*    79447    82950: contig of 3504 bp in length
*    82951    83050: gap of unknown length
*    83051    86116: contig of 3066 bp in length
*    86117    86216: gap of unknown length
*    86217    89119: contig of 2903 bp in length
*    89120    89219: gap of unknown length
*    89220    92494: contig of 3275 bp in length
*    92495    92594: gap of unknown length
*    92595    94419: contig of 1825 bp in length
*    94420    94519: gap of unknown length
*    94520    98048: contig of 3529 bp in length
*    98049    98148: gap of unknown length
*    98149   101599: contig of 3451 bp in length
*   101600   101699: gap of unknown length
*   101700   105918: contig of 4219 bp in length
*   105919   106018: gap of unknown length
*   106019   108449: contig of 2431 bp in length
*   108450   108549: gap of unknown length
*   108550   111803: contig of 3254 bp in length.
FEATURES             Location/Qualifiers

Alignment Scores:
Pred. No.:              3.58e-36        Length:         111803
Score:                  875.00          Matches:        177
Percent Similarity:     94.74%          Conservative:   3
Best Local Similarity:  93.16%          Mismatches:     10
Query Match:            45.81%          Indels:         2
DB:                     2               Gaps:           0

US-09-972-758A-2 (1-359) x AC136172 (1-111803)

Qy        170 LeuThrTrpGluGluLysLysLysPheAspGluLysGlnSerLeuArgAlaSerArgIle 189
              ||||||||| |||||||||||||||||||||||||||||||||||||||||||||:::
Db      14987 CTGACTTGGAG-GAGAAGAAAAAGTTCGACGAGAAGCAAAGCCTGCGAGCTTCGCGAGTT 14929

Qy        190 ArgAlaGluMetPheAlaLysGlyGlnProValAlaProTyrAsnThrThrGlnPheLeu 209
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      14928 CGAGCCGAGATGTC-GCCAAGGGCCAGCCGGTTGCGCCCTATAACACCACGCAGTTCNTC 14870

Qy        210 MetAspAspHisAspGlnGluGluProAspLeuLysThrGlyLeuTyrSerLysArgAla 229
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db      14869 ATGGATGACCATGATCAGGAGGAGCCTGATCTCAAAACCGGCCTTTACCCTAAGCGGGCA 14810

Qy        230 AlaAlaLysSerAspAspThrSerAspAspAspPheMetGluGluGlyGlyGluGluAsp 249
              |||||||||||||||||||||||:::|||||||:::|||||| ||||||||||||
Db      14809 GCCGCCAAATCCGACGACACCAGCGATGAGGATTTCGTGGAAGAAGCTGGTGAGGAAGAC 14750

Qy        250 GlyGlySerAspGlyMetGlyGlyAspGlySerGluPheLeuGlnArgAspPheSerGlu 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14749 GGAGGCAGCGATGGCATGGGAGGGGACGGCAGCGAGTTTCTGCAGCGGGACTTTTCCGAG 14690

Qy        270 ThrTyrGluArgTyrHisThrGluSerLeuGlnAsnMetSerLysGlnGluLeuIleLys 289
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      14689 ACTTACGAGCGGTACCACGCCGAGAGCCTGCAGAACATGAGCAAGCAGGAGCTCATCAAA 14630

Qy        290 GluTyrLeuGluLeuGluLysCysLeuSerArgMetGluAspGluAsnAsnArgLeuArg 309
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db      14629 GAGTACCTGGAGCTGGAGAAGTGCCTCTCCCGCAAGGAGGACGAAAATAACCGGCTGCGG 14570

Qy        310 LeuGluSerLysArgLeuGlyGlyAspAspAlaArgValArgGluLeuGluLeuGluLeu 329
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      14569 TTGGAAAGCAAGCGGTTGGGTGGCGTCGATGCGCGAGTGCGGGAGCTGGAGCTAGAGCTG 14510

Qy        330 AspArgLeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnGluLeuHisArgGlnGln 349
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db      14509 GACCGACTGCGCGCCGAGAACCGCCAGCTGCTGACCGAAAACGAACTGCACCGGCAGCAG 14450

Qy        350 GluArgAlaProLeuSerLysPheGlyAsp 359
              |||||||||||| ||||||||||||||||
Db      14449 GAGCGAGCGCCTCCTTCCAAGTTCGGCGAC 14420

RESULT 15
BD058159
LOCUS       BD058159                523 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Secreted expressed sequence tags (sESTs).
ACCESSION   BD058159
VERSION     BD058159.1  GI:22603765
KEYWORDS    JP 2001519666-A/14.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 523)
  AUTHORS   Jacobs,K., Mccoy,J.M., Lavallie,E.R., Racie,L.A., Merberg,D.,
            Treacy,M., Spaulding,V. and Agostino,M.J.
  TITLE     Secreted expressed sequence tags (sESTs)
  JOURNAL   Patent: JP 2001519666-A 14 23-OCT-2001;
            GENETICS INSTITUTE INC
COMMENT     PN   JP 2001519666-A/14
            PD   23-OCT-2001
            PF   10-APR-1998 JP 1998543068
            PR   10-APR-1997 US    08/835913
            PI   KENNETH JACOBS,JOHN M MCCOY,EDWARD R LAVALLIE,LISA A RACIE, PI
              DAVID MERBERG,
            PI   MAURICE TREACY,VIKKI SPAULDING,MICHAEL J AGOSTINO PC
            C12N15/12,C12N5/10,C07K14/47,C12Q1/68,A61K38/17 CC   Strandedness:
            Double;
```

```
           CC   Topology: Linear;
           FH   Key             Location/Qualifiers.
FEATURES             Location/Qualifiers
     source          1. .523
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4577"
BASE COUNT      112 a    149 c    153 g    108 t      1 others
ORIGIN
```

Alignment Scores:

Pred. No.:	5.32e-30	Length:	523
Score:	714.00	Matches:	131
Percent Similarity:	99.26%	Conservative:	3
Best Local Similarity:	97.04%	Mismatches:	1
Query Match:	37.38%	Indels:	0
DB:	6	Gaps:	0

US-09-972-758A-2 (1-359) x BD058159 (1-523)

```
Qy          1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        104 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCAGCCTCAAACTAGCAACTGTACAGGT 163

Qy         21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        164 GCTGCTGCTGTCCAGGAAGAGCTGAACCCTGAGCGCCCCCCAGGCGCGGAGGAGCGGGTG 223

Qy         41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        224 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTTCCCCCAGTTGGGTGGCCGTCCGGGG 283

Qy         61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        284 CCGGAGGGGGAAGGGAGCCTGGAATCCCAACCACCTCCCTTGCAGACCCAGGCCTGTCCA 343

Qy         81 GluSerSerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        344 GAATCTAGCTGCCTGAGAGAGGGCGAGAAGGGCCAGAATGGGGACGACTCGTCCGCTGGC 403

Qy        101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        404 GGCGACTTYCCGCCGCCGGCAGAAGTGGAACCGACGCCCGAGGCCGAGCTGCTCGCCCAG 463

Qy        121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAla 135
              ||||||||||||||||||||||:::|||:::   |||:::
Db        464 CCTTGTCATGACTCCGAGGCCAGTAAGATGGGGTCGACGGCCTCT 508
```

Search completed: February 5, 2004, 15:09:45
Job time : 4896 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 5, 2004, 13:31:10 ; Search time 92 Seconds
(without alignments)
817.046 Million cell updates/sec

Title: US-09-972-758A-2
Perfect score: 1910
Sequence: 1 MAEPFLSEYQHQPQTSNCTG.....LTENELHRQQERAPLSKFGD 359

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1910	100.0	359	9	US-09-745-763-34	Sequence 34, Appl
2	1910	100.0	359	10	US-09-972-758-2	Sequence 2, Appli
3	1910	100.0	359	15	US-10-205-823-172	Sequence 172, App
4	545.5	28.6	286	12	US-10-094-749-2862	Sequence 2862, Ap
5	134.5	7.0	546	10	US-09-976-740-44	Sequence 44, Appl
6	134.5	7.0	546	12	US-10-616-187-44	Sequence 44, Appl
7	134.5	7.0	546	14	US-10-023-529-44	Sequence 44, Appl
8	134.5	7.0	546	14	US-10-023-523-44	Sequence 44, Appl
9	130	6.8	611	9	US-09-216-393-81	Sequence 81, Appl
10	130	6.8	611	12	US-10-321-856-81	Sequence 81, Appl
11	128.5	6.7	804	12	US-10-108-260A-4161	Sequence 4161, Ap
12	128	6.7	2476	11	US-09-824-574-7	Sequence 7, Appli
13	126.5	6.6	722	12	US-10-161-051-193	Sequence 193, App
14	125	6.5	323	12	US-10-259-194A-320	Sequence 320, App
15	125	6.5	383	12	US-10-374-780A-2854	Sequence 2854, Ap
16	125	6.5	650	12	US-10-374-780A-514	Sequence 514, App
17	124	6.5	489	9	US-09-876-889-350	Sequence 350, App
18	123.5	6.5	557	9	US-09-962-055-5	Sequence 5, Appli
19	123.5	6.5	557	10	US-09-976-740-5	Sequence 5, Appli
20	123.5	6.5	557	12	US-10-616-187-5	Sequence 5, Appli
21	123.5	6.5	557	14	US-10-023-529-5	Sequence 5, Appli
22	123.5	6.5	557	14	US-10-023-523-5	Sequence 5, Appli
23	123.5	6.5	623	12	US-10-104-047-3378	Sequence 3378, Ap
24	123.5	6.5	5008	12	US-10-051-874-166	Sequence 166, App
25	121.5	6.4	342	12	US-10-317-832-118	Sequence 118, App
26	121.5	6.4	1974	12	US-10-369-493-6395	Sequence 6395, Ap
27	121	6.3	537	12	US-10-051-874-138	Sequence 138, App
28	121	6.3	538	9	US-09-827-822-9	Sequence 9, Appli
29	120	6.3	427	15	US-10-050-704-273	Sequence 273, App
30	120	6.3	483	15	US-10-050-704-272	Sequence 272, App
31	120	6.3	911	12	US-10-237-496-52	Sequence 52, Appl
32	120	6.3	911	12	US-10-242-074-52	Sequence 52, Appl
33	120	6.3	911	12	US-10-242-505-52	Sequence 52, Appl
34	120	6.3	911	12	US-10-242-574-52	Sequence 52, Appl
35	120	6.3	911	12	US-10-243-261-52	Sequence 52, Appl
36	120	6.3	911	12	US-10-243-282-52	Sequence 52, Appl
37	120	6.3	911	12	US-10-243-402-52	Sequence 52, Appl
38	120	6.3	911	12	US-10-243-431-52	Sequence 52, Appl
39	120	6.3	911	12	US-10-245-164-52	Sequence 52, Appl
40	120	6.3	911	12	US-10-244-972-52	Sequence 52, Appl
41	120	6.3	911	12	US-10-197-942-52	Sequence 52, Appl
42	120	6.3	911	12	US-10-238-196-52	Sequence 52, Appl
43	120	6.3	911	12	US-10-245-013-52	Sequence 52, Appl
44	120	6.3	911	12	US-10-295-027-260	Sequence 260, App
45	120	6.3	911	15	US-10-245-103-52	Sequence 52, Appl

ALIGNMENTS

RESULT 1
US-09-745-763-34
; Sequence 34, Application US/09745763
; Patent No. US20020065394A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; McCoy, John M.
; LaVallie, Edward R.
; Collins-Racie, Lisa A.
; Evans, Cheryl
; Merberg, David
; Treacy, Maurice
; Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; ENCODING THEM
; NUMBER OF SEQUENCES: 219
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/745,763
; FILING DATE: 18-Jun-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851

```
;   INFORMATION FOR SEQ ID NO: 34:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 359 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-09-745-763-34

  Query Match            100.0%;  Score 1910;  DB 9;  Length 359;
  Best Local Similarity  100.0%;  Pred. No. 1.2e-132;
  Matches  359;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAEPFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAEPFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPG 60

Qy         61 PEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELLAQ 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELLAQ 120

Qy        121 PCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDE 180

Qy        181 KQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDD 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDD 240

Qy        241 FMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCLSR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 FMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCLSR 300

Qy        301 MEDENNRLRLESKRLGGDDARVRELELELDRLRAENLQLLTENELHRQQERAPLSKFGD 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 MEDENNRLRLESKRLGGDDARVRELELELDRLRAENLQLLTENELHRQQERAPLSKFGD 359


RESULT 2
US-09-972-758-2
; Sequence 2, Application US/09972758
; Patent No. US20020160497A1
; GENERAL INFORMATION:
;  APPLICANT: Case Western Reserve University
;  APPLICANT:  Montano, Monica
;  APPLICANT:  Wittman, Bryan
;  TITLE OF INVENTION: Suppressors of Human Breast Cancer Cell Growth
;  FILE REFERENCE: 27708/04004
;  CURRENT APPLICATION NUMBER: US/09/972,758
;  CURRENT FILING DATE:  2001-10-05
;  PRIOR APPLICATION NUMBER: US 60/238,187
;  PRIOR FILING DATE: 2000-10-05
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
;   LENGTH: 359
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-972-758-2

  Query Match            100.0%;  Score 1910;  DB 10;  Length 359;
  Best Local Similarity  100.0%;  Pred. No. 1.2e-132;
  Matches  359;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAEPFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAEPFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPG 60

Qy         61 PEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELLAQ 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELLAQ 120

Qy        121 PCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDE 180

Qy        181 KQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDD 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDD 240

Qy        241 FMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCLSR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 FMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCLSR 300

Qy        301 MEDENNRLRLESKRLGGDDARVRELELELDRLRAENLQLLTENELHRQQERAPLSKFGD 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 MEDENNRLRLESKRLGGDDARVRELELELDRLRAENLQLLTENELHRQQERAPLSKFGD 359


RESULT 3
US-10-205-823-172
; Sequence 172, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
;  APPLICANT: Schlegel, Robert
;  APPLICANT:  Monahan, John E.
;  APPLICANT:  Endege, Wilson O.
;  APPLICANT:  Gannavarapu, Manjula
;  APPLICANT:  Gorbatcheva, Bella
;  APPLICANT:  Hoersch, Sebastian
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Wonsey, Angela M.
;  APPLICANT:  Glatt, Karen
;  APPLICANT:  Zhao, Xumei
;  APPLICANT:  Anderson, Dustin
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
;  TITLE OF INVENTION:  METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF PROSTATE CANCER
;  FILE REFERENCE: MRI-044
;  CURRENT APPLICATION NUMBER: US/10/205,823
;  CURRENT FILING DATE:  2002-07-25
;  PRIOR APPLICATION NUMBER: 60/307,982
;  PRIOR FILING DATE: 2001-07-25
;  PRIOR APPLICATION NUMBER: 60/314,356
;  PRIOR FILING DATE: 2001-08-22
;  PRIOR APPLICATION NUMBER: 60/325,020
;  PRIOR FILING DATE: 2001-09-25
;  PRIOR APPLICATION NUMBER: 60/341,746
;  PRIOR FILING DATE: 2001-12-12
;  PRIOR APPLICATION NUMBER: 60/362,158
;  PRIOR FILING DATE: 2002-03-05
;  NUMBER OF SEQ ID NOS: 455
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 172
;   LENGTH: 359
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-205-823-172

  Query Match            100.0%;  Score 1910;  DB 15;  Length 359;
  Best Local Similarity  100.0%;  Pred. No. 1.2e-132;
  Matches  359;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAEPFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAEPFLSEYQHQPQTSNCTGAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPG 60

Qy         61 PEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELLAQ 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PEGEGSLESQPPPLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELLAQ 120

Qy        121 PCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDE 180
```

```
Qy        181 KQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDD 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDD 240

Qy        241 FMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCLSR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 FMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCLSR 300

Qy        301 MEDENNRLRLESKRLGGDDARVRELELELDRLRAENLQLLTENELHRQQERAPLSKFGD 359
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 MEDENNRLRLESKRLGGDDARVRELELELDRLRAENLQLLTENELHRQQERAPLSKFGD 359


RESULT 4
US-10-094-749-2862
; Sequence 2862, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
;  APPLICANT: ISOGAI, TAKAO
;  APPLICANT:  SUGIYAMA, TOMOYASU
;  APPLICANT:  OTSUKI, TETSUJI
;  APPLICANT:  WAKAMATSU, AI
;  APPLICANT:  SATO, HIROYUKI
;  APPLICANT:  ISHII, SHIZUKO
;  APPLICANT:  YAMAMOTO, JUN-ICHI
;  APPLICANT:  ISONO, YUUKO
;  APPLICANT:  HIO, YURI
;  APPLICANT:  OTSUKA, KAORU
;  APPLICANT:  NAGAI, KEIICHI
;  APPLICANT:  IRIE, RYOTARO
;  APPLICANT:  TAMECHIKA, ICHIRO
;  APPLICANT:  SEKI, NAOHIKO
;  APPLICANT:  YOSHIKAWA, TSUTOMU
;  APPLICANT:  OTSUKA, MOTOYUKI
;  APPLICANT:  NAGAHARI, KENJI
;  APPLICANT:  MASUHO, YASUHIKO
;  TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
;  FILE REFERENCE: 084335/0160
;  CURRENT APPLICATION NUMBER: US/10/094,749
;  CURRENT FILING DATE:  2002-03-12
;  PRIOR APPLICATION NUMBER: 60/350,435
;  PRIOR FILING DATE: 2002-01-24
;  PRIOR APPLICATION NUMBER: JP 2001-328381
;  PRIOR FILING DATE: 2001-09-14
;  NUMBER OF SEQ ID NOS: 3381
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2862
;   LENGTH: 286
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-094-749-2862

  Query Match             28.6%;  Score 545.5;  DB 12;  Length 286;
  Best Local Similarity   43.8%;  Pred. No. 2.3e-32;
  Matches  128;  Conservative   36;  Mismatches   93;  Indels   35;  Gaps    6;

Qy         73 PLQTQACPESSCLREGEKGQNGDDS--------SAGGDFPPPAEVEPTPEAELLAQPC-- 122
              | ||    ||    |  |       |         :||  |   :|   | | ||
Db          5 PNQTACNAESPVALEEAKTSGAPGSPQTPPERHDSGGSLPLTPRMESHSEDEDLAGAVGG 64

Qy        123 --HDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKKFDE 180
                 :| : : ´:|      |         | :||||||||:||||:|| :|:| ||:: ||
Db         65 LGWNSRSPRTQSPGGCSAE------AVLARKKHRRRPSKRKRHWRPYLELSWAEKQQRDE 118

Qy        181 KQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEP--DLKTGLYSKRAAAKSDDTSD 238
              :|| ||||:| ||||||||||||||||||:| | ||| |:  |:    :: :|
Db        119 RQSQRASRVREEMFAKGQPVAPYNTTQFLMNDRDPEEPNLDVPHGISHPGSSGES----- 173

Qy        239 DDFMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKEYLELEKCL 298
                    | | ||| |    || ::||||||||:||||||  |||||:::|||||| |
Db        174 ---------EAGDSDGRGRAHGEFQRKDFSETYERFHTESLQGRSKQELVRDYLELEKRL 224

Qy        299 SRMEDENNRLRLESKRLGGDDAR-VRELELELDRLRAENLQLLTENELHRQQ 349
                |: |:|  ||:      |    | | ||  |: ||| || :|  ||::  ::
Db        225 SQAEEETRRLQQLQACTGQQSCRQVEELAAEVQRLRTENQRLRQENQMWNRE 276


RESULT 5
US-09-976-740-44
; Sequence 44, Application US/09976740
; Publication No. US20020194633A1
; GENERAL INFORMATION:
;  APPLICANT: Lees, Ann M.
;  APPLICANT:  Lees, Robert S.
;  APPLICANT:  Law, Simon W.
;  APPLICANT:  Arjona, Anibal A.
;  TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
;  TITLE OF INVENTION:  PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
;  TITLE OF INVENTION:  ATHEROSCLEROSIS
;  FILE REFERENCE: 10797-004001
;  CURRENT APPLICATION NUMBER: US/09/976,740
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: 09/616,289
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: US 08/979,608
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 60/031,930
;  PRIOR FILING DATE: 1996-11-27
;  PRIOR APPLICATION NUMBER: US 60/048,547
;  PRIOR FILING DATE: 1997-06-03
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44
;   LENGTH: 546
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-976-740-44

  Query Match             7.0%;  Score 134.5;  DB 10;  Length 546;
  Best Local Similarity   21.9%;  Pred. No. 0.087;
  Matches   86;  Conservative   51;  Mismatches  134;  Indels  121;  Gaps   14;

Qy         20 GAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPGPEGEGSLESQPPPLQTQAC 79
              |||   :: ||:  ||  |  ||      |:| | :      | ||   || |
Db          9 GAA---KQSNPKSSPGQPEAGPEGAQERPSQAAPAV------EAEGPGSSQAP-----RK 54

Qy         80 PESSCLREGEKGQNGD-----------------DSSAGGDFPPPAEVEPTPEAELLAQP 121
              || :  |  : |  |                    |:: ||     |: ||       |:|
Db         55 PEGAQARTAQSGALRDVSEELSRQLEDILSTYCVDNNQGGPGEDGAQGEP-------AEP 107

Qy        122 CHDSEASKL---------GAPAAGGEEEWG---------QQQRQLGKKKHRRRPSKKKRH 163
                |:| |:          |   ||:|           :|  ::| : |||   |||
Db        108 -EDAEKSRTYVARNGEPEPTPVVNGEKEPSKGDPNTEEIRQSDEVGDRDHRRPQEKKKAK 166

Qy        164 W---------------KPYYKLTWEEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNTTQ 207
                              |  ||    ||  :  :  | |: : ::  | |       :|
Db        167 GLGKEITLLMQTLNTLSTPEEKLAALCKKYAELLEEHRNSQKQMKLLQKKQ-------SQ 219

Qy        208 FLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDF 267
               :      :| |   | :||   |:|                              | | |:
Db        220 LV------QEKDHLRGEHSKAVLARSK------------------------LESLCREL 248

Qy        268 SETYERYHTESLQNMSKQELIKEYLELEKCLSRMEDENNRLRLESKRLGGDDARVRELEL 327
                  |  ||:    |   :|  : ::  |  :   | ::|: ::    ::::|:  :
Db        249 QR-----HNRSLKEEGVQRAREEEEKRKEVTSHFQVTLNDIQLQMEQHNERNSKLRQENM 303

Qy        328 ELDRLRAENLQLLTENELHRQQERAPLSKFGD 359
              ||    || |: | |    |::    : |  |
Db        304 EL----AERLKKLIEQYELREEHIDKVFKHKD 331


RESULT 6
US-10-616-187-44
; Sequence 44, Application US/10616187
; Publication No. US20040013668A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Lees, Ann M.
;  APPLICANT:  Lees, Robert S.
;  APPLICANT:  Law, Simon W.
;  APPLICANT:  Arjona, Anibal A.
;  TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
;  TITLE OF INVENTION:  PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
;  TITLE OF INVENTION:  ATHEROSCLEROSIS
;  FILE REFERENCE: 10797-004001
;  CURRENT APPLICATION NUMBER: US/10/616,187
;  CURRENT FILING DATE:  2003-07-09
;  PRIOR APPLICATION NUMBER: US/09/616,289
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: US 09/517,849
;  PRIOR FILING DATE: 2000-03-02
;  PRIOR APPLICATION NUMBER: US 08/979,608
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 60/031,930
;  PRIOR FILING DATE: 1996-11-27
;  PRIOR APPLICATION NUMBER: US 60/048,547
;  PRIOR FILING DATE: 1997-06-03
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44
;   LENGTH: 546
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-616-187-44

  Query Match             7.0%;  Score 134.5;  DB 12;  Length 546;
  Best Local Similarity   21.9%;  Pred. No. 0.087;
  Matches   86;  Conservative   51;  Mismatches  134;  Indels  121;  Gaps   14;

Qy         20 GAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPGPEGEGSLESQPPPLQTQAC 79
              |||   :: ||:  ||  |  ||       |:| | :      | ||   || |
Db          9 GAA---KQSNPKSSPGQPEAGPEGAQERPSQAAPAV------EAEGPGSSQAP-----RK 54

Qy         80 PESSCLREGEKGQNGD-----------------DSSAGGDFPPPAEVEPTPEAELLAQP 121
              || :  |  : |  |                 |:: ||     |: ||       |:|
Db         55 PEGAQARTAQSGALRDVSEELSRQLEDILSTYCVDNNQGGPGEDGAQGEP-------AEP 107

Qy        122 CHDSEASKL---------GAPAAGGEEEWG---------QQQRQLGKKKHRRRPSKKKRH 163
                |:| |:           |   ||:|         :|  ::| : |||   |||
Db        108 -EDAEKSRTYVARNGEPEPTPVVNGEKEPSKGDPNTEEIRQSDEVGDRDHRRPQEKKKAK 166

Qy        164 W----------------KPYYKLTWEEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNTTQ 207
                              |  ||    ||  :  :  | |: : ::  | |       :|
Db        167 GLGKEITLLMQTLNTLSTPEEKLAALCKKYAELLEEHRNSQKQMKLLQKKQ-------SQ 219

Qy        208 FLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDF 267
               :      :| |   | :||   |:|                              | | |:
Db        220 LV------QEKDHLRGEHSKAVLARSK-----------------------LESLCREL 248

Qy        268 SETYERYHTESLQNMSKQELIKEYLELEKCLSRMEDENNRLRLESKRLGGDDARVRELEL 327
                  |  ||:    |   :|  : ::  |  :   | ::|: ::    ::::|:  :
Db        249 QR-----HNRSLKEEGVQRAREEEEKRKEVTSHFQVTLNDIQLQMEQHNERNSKLRQENM 303

Qy        328 ELDRLRAENLQLLTENELHRQQERAPLSKFGD 359
              ||    || |: | |    |::    : |  |
Db        304 EL----AERLKKLIEQYELREEHIDKVFKHKD 331


RESULT 7
US-10-023-529-44
; Sequence 44, Application US/10023529
; Publication No. US20020129388A1
; GENERAL INFORMATION:
;  APPLICANT: Lees, Ann M.
;  APPLICANT:  Lees, Robert S.
;  APPLICANT:  Law, Simon W.
;  APPLICANT:  Arjona, Anibal A.
;  TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
;  TITLE OF INVENTION:  PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
;  TITLE OF INVENTION:  ATHEROSCLEROSIS
;  FILE REFERENCE: 10797-004001
;  CURRENT APPLICATION NUMBER: US/10/023,529
;  CURRENT FILING DATE:  2001-12-17
;  PRIOR APPLICATION NUMBER: 09/616,289
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: US 09/517,849
;  PRIOR FILING DATE: 2000-03-02
;  PRIOR APPLICATION NUMBER: US 08/979,608
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 60/031,930
;  PRIOR FILING DATE: 1996-11-27
;  PRIOR APPLICATION NUMBER: US 60/048,547
;  PRIOR FILING DATE: 1997-06-03
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44
;   LENGTH: 546
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-023-529-44

  Query Match             7.0%;  Score 134.5;  DB 14;  Length 546;
  Best Local Similarity   21.9%;  Pred. No. 0.087;
  Matches   86;  Conservative   51;  Mismatches  134;  Indels  121;  Gaps   14;

Qy         20 GAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPGPEGEGSLESQPPPLQTQAC 79
              |||   :: ||:  ||  |  ||       |:| | :      | ||   || |
Db          9 GAA---KQSNPKSSPGQPEAGPEGAQERPSQAAPAV------EAEGPGSSQAP-----RK 54

Qy         80 PESSCLREGEKGQNGD-----------------DSSAGGDFPPPAEVEPTPEAELLAQP 121
              || :  |  : |  |                 |:: ||     |: ||       |:|
Db         55 PEGAQARTAQSGALRDVSEELSRQLEDILSTYCVDNNQGGPGEDGAQGEP-------AEP 107

Qy        122 CHDSEASKL---------GAPAAGGEEEWG---------QQQRQLGKKKHRRRPSKKKRH 163
                |:| |:           |   ||:|         :|  ::| : |||   |||
Db        108 -EDAEKSRTYVARNGEPEPTPVVNGEKEPSKGDPNTEEIRQSDEVGDRDHRRPQEKKKAK 166

Qy        164 W----------------KPYYKLTWEEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNTTQ 207
                              |  ||    ||  :  :  | |: : ::  | |       :|
Db        167 GLGKEITLLMQTLNTLSTPEEKLAALCKKYAELLEEHRNSQKQMKLLQKKQ-------SQ 219

Qy        208 FLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDF 267
               :      :| |   | :||   |:|                              | | |:
Db        220 LV------QEKDHLRGEHSKAVLARSK-----------------------LESLCREL 248

Qy        268 SETYERYHTESLQNMSKQELIKEYLELEKCLSRMEDENNRLRLESKRLGGDDARVRELEL 327
                  |  ||:    |   :|  : ::  |  :   | ::|: ::    ::::|:  :
Db        249 QR-----HNRSLKEEGVQRAREEEEKRKEVTSHFQVTLNDIQLQMEQHNERNSKLRQENM 303

Qy        328 ELDRLRAENLQLLTENELHRQQERAPLSKFGD 359
              ||    || |: | |    |::    : |  |
Db        304 EL----AERLKKLIEQYELREEHIDKVFKHKD 331


RESULT 8
US-10-023-523-44
; Sequence 44, Application US/10023523
; Publication No. US20020152485A1
; GENERAL INFORMATION:
;  APPLICANT: Lees, Ann M.
;  APPLICANT:  Lees, Robert S.
;  APPLICANT:  Law, Simon W.
;  APPLICANT:  Arjona, Anibal A.
;  TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
;  TITLE OF INVENTION:  PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
;  TITLE OF INVENTION:  ATHEROSCLEROSIS
;  FILE REFERENCE: 10797-004001
;  CURRENT APPLICATION NUMBER: US/10/023,523
;  CURRENT FILING DATE:  2001-12-17
;  PRIOR APPLICATION NUMBER: US/09/616,289
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: US 09/517,849
;  PRIOR FILING DATE: 2000-03-02
```

```
;  PRIOR APPLICATION NUMBER: US 08/979,608
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 60/031,930
;  PRIOR FILING DATE: 1996-11-27
;  PRIOR APPLICATION NUMBER: US 60/048,547
;  PRIOR FILING DATE: 1997-06-03
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44
;   LENGTH: 546
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-023-523-44

  Query Match             7.0%;  Score 134.5;  DB 14;  Length 546;
  Best Local Similarity   21.9%;  Pred. No. 0.087;
  Matches   86;  Conservative   51;  Mismatches  134;  Indels  121;  Gaps   14;

Qy         20 GAAAVQEELNPERPPGAEERVPEEDSRWQSRAFPQLGGRPGPEGEGSLESQPPPLQTQAC 79
              |||   :: ||:  ||  |  ||       |:| | :      | ||   || |
Db          9 GAA---KQSNPKSSPGQPEAGPEGAQERPSQAAPAV------EAEGPGSSQAP-----RK 54

Qy         80 PESSCLREGEKGQNGD-----------------DSSAGGDFPPPAEVEPTPEAELLAQP 121
              || :  |  : |   |                   |:: ||     |: ||       |:|
Db         55 PEGAQARTAQSGALRDVSEELSRQLEDILSTYCVDNNQGSPGEDGAQGEP-------AEP 107

Qy        122 CHDSEASKL---------GAPAAGGEEEWG---------QQQRQLGKKKHRRRPSKKKRH 163
               |:| |:           |   ||:|          :|  ::| : |||   |||
Db        108 -EDAEKSRTYVARNGEPEPTPVVNGEKEPSKGDPNTEEIRQSDEVGDRDHRRPQEKKKAK 166

Qy        164 W----------------KPYYKLTWEEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNTTQ 207
                             |  ||     ||  :  :  | |: : ::  | |        :|
Db        167 GLGKEITLLMQTLNTLSTPEEKLAALCKKYAELLEEHRNSQKQMKLLQKKQ-------SQ 219

Qy        208 FLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDF 267
               :     :| |   | :||   |:|                              | | |:
Db        220 LV------QEKDHLRGEHSKAVLARSK------------------------LESLCREL 248

Qy        268 SETYERYHTESLQNMSKQELIKEYLELEKCLSRMEDENNRLRLESKRLGGDDARVRELEL 327
                  |  ||:    |   :|  : ::  |  :   | ::|: ::    ::::|:  :
Db        249 QR-----HNRSLKEEGVQRAREEEEKRKEVTSHFQVTLNDIQLQMEQHNERNSKLRQENM 303

Qy        328 ELDRLRAENLQLLTENELHRQQERAPLSKFGD 359
              ||    || |: | |    |::   : |  |
Db        304 EL----AERLKKLIEQYELREEHIDKVFKHKD 331


RESULT 9
US-09-216-393-81
; Sequence 81, Application US/09216393
; Patent No. US20010014447A1
; GENERAL INFORMATION:
;  APPLICANT: Milhausen, Michael James
;  TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
;  TITLE OF INVENTION: USES THEREOF
;  FILE REFERENCE: TX-1-C2
;  CURRENT APPLICATION NUMBER: US/09/216,393
;  CURRENT FILING DATE: 1998-12-18
;  EARLIER APPLICATION NUMBER: 08/994,825
;  EARLIER FILING DATE: 1997-12-19
;  NUMBER OF SEQ ID NOS: 364
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 81
;   LENGTH: 611
;   TYPE: PRT
;   ORGANISM: Toxoplasma gondii
US-09-216-393-81

  Query Match             6.8%;  Score 130;  DB 9;  Length 611;
  Best Local Similarity   22.2%;  Pred. No. 0.21;
  Matches   82;  Conservative   58;  Mismatches  155;  Indels   74;  Gaps   15;

Qy         13 PQTSNCTGAAAVQEELN-PERPPGAEERVP-----EEDSRWQSRAFP----QLGGRPGPE 62
              |: |       : |  |: ||:  |:     |      :  : : | | |    :|   ||
Db        189 PRFSKSDVCCSPQARLSLPEQSLGSSPSSPISVTNDVYALFDSSASPLHAGELSSLPGAV 248

Qy         63 GEGSLESQPP----PLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELL 118
                      |    |  : ||   ||                    | |: || :
Db        249 SASERLLTAPAEIGPSASSACLSVSC-------------------GPGEMSPTADTTR- 287

Qy        119 AQPCHDSEASKLGAPAAGGE----EEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEE 174
                 ||:|  :          |    ||  :::|:: ::| | |  :::|      :   ||
Db        288 ----HDAEERERRRAEEEKERERQEEEERERRRVEEEKERERQEEEERE----RRRVEEE 339

Qy        175 K---KKFDEKQSLRA----SRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSK 227
              |   :: ||::  |      :| | |   : : |        |       ::|| : :     |
Db        340 KARQREEDERERRRVEEEKARQREEEERERRRVEEEKARQ----REEEEERERRRVEEEK 395

Qy        228 RAAAKSDDTSDDDFMEE----GGEEDGGSDGMGGDGSEFLQRDFSETYE--RYHTESLQN 281
                : ::  :   :||       ||:  :|   :  :  ||:  |  |  |   | :
Db        396 ARQREEEEERERRRVEEEKARQREEEEEREGRRVEEEKARQREEEEEREGRRVEEEKARQ 455

Qy        282 MSKQELIKEYLELEKCLSRMEDENNRLRLESKRLGGDDARVRELELELDRLRAENLQLLT 341
               ::|  :  :| ||    | |:|  | |:| :: |:| | | :| | | |
Db        456 REEEERERRRVEEEKERERQEEERERRRVEEEK-----ERERQEEEERERRRVEE----- 505

Qy        342 ENELHRQQE 350
              | |  ||:|
Db        506 EKERERQEE 514


RESULT 10
US-10-321-856-81
; Sequence 81, Application US/10321856
; Publication No. US20030194393A1
; GENERAL INFORMATION:
;  APPLICANT: Milhausen, Michael James
;  TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND USES THERE
;  FILE REFERENCE: TX-1-C2-1
;  CURRENT APPLICATION NUMBER: US/10/321,856
;  CURRENT FILING DATE:  2002-12-17
;  PRIOR APPLICATION NUMBER: 09/216,393
;  PRIOR FILING DATE: 1998-12-18
;  PRIOR APPLICATION NUMBER: 08/994,825
;  PRIOR FILING DATE: 1997-12-19
;  NUMBER OF SEQ ID NOS: 366
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 81
;   LENGTH: 611
;   TYPE: PRT
;   ORGANISM: Toxoplasma gondii
US-10-321-856-81

  Query Match             6.8%;  Score 130;  DB 12;  Length 611;
  Best Local Similarity   22.2%;  Pred. No. 0.21;
  Matches   82;  Conservative   58;  Mismatches  155;  Indels   74;  Gaps   15;

Qy         13 PQTSNCTGAAAVQEELN-PERPPGAEERVP-----EEDSRWQSRAFP----QLGGRPGPE 62
              |: |       : |  |: ||:  |:     |      :  : : | | |    :|   ||
Db        189 PRFSKSDVCCSPQARLSLPEQSLGSSPSSPISVTNDVYALFDSSASPLHAGELSSLPGAV 248

Qy         63 GEGSLESQPP----PLQTQACPESSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELL 118
                      |    |  : ||   ||                    | |: || :
Db        249 SASERLLTAPAEIGPSASSACLSVSC-------------------GPGEMSPTADTTR- 287

Qy        119 AQPCHDSEASKLGAPAAGGE----EEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEE 174
                 ||:|  :          |    ||  :::|:: ::| | |  :::|      :   ||
Db        288 ----HDAEERERRRAEEEKERERQEEEERERRRVEEEKERERQEEEERE----RRRVEEE 339

Qy        175 K---KKFDEKQSLRA----SRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSK 227
              |   :: ||::  |      :| | |   : : |        |       ::|| : :     |
Db        340 KARQREEDERERRRVEEEKARQREEEERERRRVEEEKARQ----REEEEERERRRVEEEK 395

Qy        228 RAAAKSDDTSDDDFMEE----GGEEDGGSDGMGGDGSEFLQRDFSETYE--RYHTESLQN 281
                : ::  :   :||       ||:  :|   :  :  ||:  |  |  |   | :
Db        396 ARQREEEEERERRRVEEEKARQREEEEEREGRRVEEEKARQREEEEEREGRRVEEEKARQ 455
```

```
Qy        282 MSKQELIKEYLELEKCLSRMEDENNRLRLESKRLGGDDARVRELELELDRLRAENLQLLT 341
Db        456 REEEERERRRVEEEKERERQEEERERRRVEEEK-----ERERQEEEERERRRVEE----- 505

Qy        342 ENELHRQQE 350
Db        506 EKERERQEE 514
```

RESULT 11
US-10-108-260A-4161
; Sequence 4161, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4161
; LENGTH: 804
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-4161

Query Match 6.7%; Score 128.5; DB 12; Length 804;
Best Local Similarity 21.3%; Pred. No. 0.38;
Matches 96; Conservative 52; Mismatches 137; Indels 165; Gaps 21;

```
Qy          5 FLSEYQHQPQTSNCTGAAAVQEELNPERPPGAE--ERV---PEE----------DSRWQ 48
Db        173 FSSIRRH--RKSKVTGA--------EQSEPGAKGPERVRARPHEHVSSAPQVPCFEETFQ 222

Qy         49 S----RAFPQLGGRPGPEGEGSLESQPPPLQTQAC--------------------PESSC 84
Db        223 APRKENANPQ--DAPGPKVSPTPEPSPPATEKMACKDPEKPMEACASAHVQPKPAPEASS 280

Qy         85 LREGEKGQNGDDSSAGGDFPPPAEV-EP----------------TPEAELLAQPCHDSE 126
Db        281 LEEPHSPETGEKVVAGEVNPPNGPVGDPLSLLFGDVTSLKSFDSLTGCGDIIAEQDMDSM 340

Qy        127 ASKLGAPAAGG------------------------------EEEWGQQQRQLGKKKHRR 155
Db        341 TDSM---ASGGQRANRDGTKRSSCLVTYQGGGEEMALPDDDDEEEEEEEEVELEEEE--- 394

Qy        156 RPSKKKRHWKPYYKLTWEEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNTTQ------FL 209
Db        395 ----------------EEVKEEEEDDDLEYLWETAQMYPRPNMNLGYHPTTSPGHHGYM 437

Qy        210 MDDHDQEEPDLKTG--LYSKRAAAKSDDTSDDDFMEE--GGEEDGGSDGMGGDGSEFLQR 265
Db        438 LLDPVRSYPGLAPGELLTPQSDQQESAPNSDEGYYDSTTPGFEDDSGEALGLVRRDCLPR 497

Qy        266 DF---SETYERYH-TESLQNM-----------SKQELIKEYLELEKCLS-----RMEDE 304
Db        498 DSYSGDALYEFYEPDDSLENSPPGDDCLYDLHGRSSEMFDPFLNFEPFLSSRPPGAMETE 557

Qy        305 NNRL----------RLESKRLGGDDARVRE 324
Db        558 EERLVTIQKQLLYWELRREQLEAQEARARE 587
```

RESULT 12
US-09-824-574-7
; Sequence 7, Application US/09824574
; Publication No. US20030077800A1
; GENERAL INFORMATION:
; APPLICANT: Rouleau, Natalie
; APPLICANT: Moilanen, Anu-Maarit
; APPLICANT: Palvimo, Jorma J.
; APPLICANT: Jnne, Olli A,
; TITLE OF INVENTION: ARIP4 Gene and Protein
; FILE REFERENCE: 2630-109
; CURRENT APPLICATION NUMBER: US/09/824,574
; CURRENT FILING DATE: 2001-04-03
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 2476
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-824-574-7

Query Match 6.7%; Score 128; DB 11; Length 2476;
Best Local Similarity 19.7%; Pred. No. 1.6;
Matches 78; Conservative 52; Mismatches 130; Indels 136; Gaps 13;

```
Qy         43 EDSRWQSRAFPQLGGRPGPEGE---GSLESQPPPLQTQACPESSCLREGEKGQNGDDSSA 99
Db       1142 EDNKKQKKQRTSAKKKTGNTKEKKRNSLRATPKRKQVDITSSSSDIGDDDQNSAGEESSD 1201

Qy        100 GGDFPP--------------------------PAEVEP-----TPE---------AELLA 119
Db       1202 EQKIKPVTENLVLPSHTGFCQSSGDEALSKSVPATVDDDDDDNDPENRIAKKMLLEEIKA 1261

Qy        120 QPCHDSEASKLGAPAAGGEEEWGQQQ----------------RQLGKKKHRRRPSKKK- 161
Db       1262 NLSSDEDGSSDDEPDGGGKKRIGKQSEESPADDGELRREQLAVNQVNSESDSDSEESKKP 1321

Qy        162 --RHWKPYYKLTWEEKKKFDEKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPD 219
Db       1322 RYRHRLLRHKLTLSDGESGEEK----PTKPKEHKEAKGR---------------------- 1356

Qy        220 LKTGLYSKRAAAKSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSETYERYHTESL 279
Db       1357 -------NRRKVSSEDSEDTDFQESGVSEEVSE-------SEDEQRPRTRSAKKAELEEN 1402

Qy        280 QNMSKQELIKEYLELEKCLSR----------------------MEDENNRLRLESKRL 315
Db       1403 QRSYKQKKKRRRIKVQEDSSSENKSHSEEDKKEGDEEDEEDEDEDEEDEND----DSKSP 1458

Qy        316 GGDDARVRELELELDRLRAENLQLLTENELHRQQER 351
Db       1459 GKGRKKIRKI-LKDDKLRTE-----TQNALKEEEER 1488
```

RESULT 13
US-10-161-051-193
; Sequence 193, Application US/10161051
; Publication No. US20030152945A1
; GENERAL INFORMATION:
; APPLICANT: Peter Deak
; APPLICANT: David M Glover
; APPLICANT: Carol Midgley
; TITLE OF INVENTION: Cell cycle progression proteins
; FILE REFERENCE: CCI-021CP
; CURRENT APPLICATION NUMBER: US/10/161,051
; CURRENT FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: GB 0007268.6
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 193
; LENGTH: 722
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-161-051-193

Query Match 6.6%; Score 126.5; DB 12; Length 722;
Best Local Similarity 22.5%; Pred. No. 0.47;
Matches 84; Conservative 47; Mismatches 149; Indels 93; Gaps 16;

```
Qy         12 QPQTSNC--TGAAAVQEELNPERPPGAEERVP-EEDSRWQSRAFPQLGGRPGPEGEGSLE 68
```

```
              :|  | |   |    :|:   |     ||  | ||| : :|      |   |  | ||
Db        324 EPDESLCLEQGDEEEEEDAEDEDLDEDEEDPPSEEDKKRKS------GKSSGGAGRGSAR 377

Qy         69 SQP--PPLQT-----QACPESSCLREGE------KGQNGDDSSAGGDFPPPAEVEPTPEA 115
              :   |   |      |  : |   : |      | :  ||| :| |: | |
Db        378 NSTGRPRRATAGKKMSAYVDFSSSDDSEQKVAVPKRRRNDDSESGSDYNPSAN------- 430

Qy        116 ELLAQPCHDSEASKLGAPAAGGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEK 175
                    ||:  : |   | | :   :  |    :| ||| |  :           ||:
Db        431 -------SDSDGGRGGGAGAAGRKVPSRGGRGRPARKSRRRNSDSEE----------EEE 473

Qy        176 KKFDEKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEEPDLKTGLYSK-RAAA--- 231
               :  :  |    | |  :  :|:| ||                    |   : | ||
Db        474 SEVSDADSDVPKRKRGSVGKRGRPAAP----------------ASAGRRGRGRGAASRK 516

Qy        232 KSDDTSDDDFMEEGGEEDGGSDGMGGDGSEFLQRDFSET------YERYHTESLQNMSKQ 285
               : |  |:|: : |  ||:  ||    | ||  : :   :      |     |     ||:
Db        517 RKDSDSEDEEVSEDEEEEDVSD-FASDQSEVCKFNLISSIWCFIKYMPIFQEERPKKSKK 575

Qy        286 ELIKEYLELEKCLSRMEDENNRLRLESKRLGGDDARVRELELELDRLRAENLQLLTENEL 345
               :                :|::   :||  |  |:| :: : |      |:  :  ::|
Db        576 PITPA------------KNSKANNKSKPAGKADSRSKKSKKESSE---EDDDVDDKDE- 618

Qy        346 HRQQERAPLSKFG 358
                  |  ||:| |
Db        619 --SDEDEPLTKKG 629


RESULT 14
US-10-259-194A-320
; Sequence 320, Application US/10259194A
; Publication No. US20040010815A1
; GENERAL INFORMATION:
;  APPLICANT: Lange, Markus B.
;  APPLICANT:  Ghassemian, Majid
;  APPLICANT:  Briggs, Steven P.
;  APPLICANT:  Cooper, Bret
;  APPLICANT:  Glazebrook, Jane
;  APPLICANT:  Goff, Stephen A.
;  APPLICANT:  Katagiri, Fumiyaki
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Moughamer, Todd
;  APPLICANT:  Provart, Nicholas
;  APPLICANT:  Ricke, Darrell
;  APPLICANT:  Zhu, Tong
;  TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES
;  FILE REFERENCE: 70029-NP
;  CURRENT APPLICATION NUMBER: US/10/259,194A
;  CURRENT FILING DATE:  2003-01-07
;  PRIOR APPLICATION NUMBER: US 60/325,277
;  PRIOR FILING DATE: 2001-09-26
;  PRIOR APPLICATION NUMBER: US 60/370,743
;  PRIOR FILING DATE: 2002-04-04
;  PRIOR APPLICATION NUMBER: US 60/370,620
;  PRIOR FILING DATE: 2002-04-04
;  NUMBER OF SEQ ID NOS: 662
;  SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 320
;   LENGTH: 323
;   TYPE: PRT
;   ORGANISM: Oryza sativa
US-10-259-194A-320

  Query Match             6.5%;  Score 125;  DB 12;  Length 323;
  Best Local Similarity   21.6%;  Pred. No. 0.23;
  Matches   66;  Conservative   60;  Mismatches  130;  Indels   50;  Gaps   13;

Qy         64 EGSLESQPPPLQTQACPESSCLREGEKG-QNGDDSSAG----GDFPPPAEVEPTPEAELL 118
              : ||:     | :|   :||   || :  :|   :::      ||           |:|
Db          7 QNSLKGTDKHLPSQVRTDSSGEEEGSRARENARKANSSRRKIKDF----------SADLE 56

Qy        119 AQPCHDSEASKLGAPAA-GGEEEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEEKKK 177
               :  ||  :    :|:   |:|   :|  || :    |  :: :|       | :
Db         57 LKKAHDDLSPGEKSPSQHSGKETRRKQNNQLSESSEDERDGRRTKH-------TVDSPDD 109

Qy        178 FDEKQSLRASRIRAEMFAKGQPVAPYNTTQFLMDDHDQEE--PDLKTGLYSKRAAAKSDD 235
                :||    ||            |  :|:  :  |  : |:    |: |: ||:  || ||
Db        110 SQQKQH-TPSR-----------VGMHNS--YSKDGMNSEDAIKGLRDGMASKKYPAKIDD 155

Qy        236 TSDDD----FMEEGGEEDGGSD-GMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKE 290
               |: :    | ::  :  | ::   |  |||   :  | : :| | :| ::  : :   :
Db        156 DSESEDGSPFRKDKRKAHGNNNIDSGSSGSEESGKHRSHSEKRKHKKSRKHKRRYDDSSD 215

Qy        291 YLELE---KCLSRMEDENNRLRLESKRLGGDDARVRELELELDRLR---AENLQLLTENE 344
                  |   |   |   |  ||| | :|   ::   |: : :  :|:   || : : :: |
Db        216 ESNSESDDKESKRRRKEEKRLRKEERRRRREERHRRKADRQASKLKLKHAETVDMASDLE 275

Qy        345 LHRQQE 350
                |: :
Db        276 KDRESD 281


RESULT 15
US-10-374-780A-2854
; Sequence 2854, Application US/10374780A
; Publication No. US20040019927A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, Bradley K
;  APPLICANT:  Riechmann, Jose Luis
;  APPLICANT:  Jiang, Cai-Zhong
;  APPLICANT:  Heard, Jacqueline E
;  APPLICANT:  Haake, Volker
;  APPLICANT:  Creelman, Robert A
;  APPLICANT:  Ratcliffe, Oliver
;  APPLICANT:  Adam, Luc J
;  APPLICANT:  Reuber, T. Lynne
;  APPLICANT:  Keddie, James
;  APPLICANT:  Broun, Pierre E
;  APPLICANT:  Pilgrim, Marsha L
;  APPLICANT:  Dubell III, Arnold T
;  APPLICANT:  Pineda, Omaira
;  APPLICANT:  Yu, Guo-Liang
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
;  FILE REFERENCE: MBI-0047 CIP
;  CURRENT APPLICATION NUMBER: US/10/374,780A
;  CURRENT FILING DATE:  2003-02-25
;  PRIOR APPLICATION NUMBER: 09/837,944
;  PRIOR FILING DATE: 2001-04-18
;  PRIOR APPLICATION NUMBER: 60/310,847
;  PRIOR FILING DATE: 2001-08-09
;  PRIOR APPLICATION NUMBER: 09/934,455
;  PRIOR FILING DATE: 2001-08-22
;  PRIOR APPLICATION NUMBER: 60/336,049
;  PRIOR FILING DATE: 2001-11-19
;  PRIOR APPLICATION NUMBER: 60/338,692
;  PRIOR FILING DATE: 2001-12-11
;  PRIOR APPLICATION NUMBER: 10/171,468
;  PRIOR FILING DATE: 2002-06-14
;  PRIOR APPLICATION NUMBER: 10/225,066
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: 10/225,067
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: 10/225,068
;  PRIOR FILING DATE: 2002-08-09
;  NUMBER OF SEQ ID NOS: 2906
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2854
;   LENGTH: 383
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   OTHER INFORMATION: G2373
US-10-374-780A-2854

  Query Match             6.5%;  Score 125;  DB 12;  Length 383;
  Best Local Similarity   20.6%;  Pred. No. 0.29;
```

Matches 84; Conservative 42; Mismatches 116; Indels 166; Gaps 18;

```
Qy         28 LNPERPPGAEERVPEEDSRWQSRAFPQLGGRPGPEGEGSLESQPPPLQTQA----CPE-- 81
              |:|: || :   :|  |         ||    ||      | || | :   |     |
Db         18 LSPQAPP-SPPILPTNDVTVAVVKKPQ----PG------LSSQSPSMNALALVVHTPSVT 66

Qy         82 ---SSCLREGEKGQNGDDSSAGGDFPPPAEVEPTPEAELLAQPCHDSEASKLGAPAAGGE 138
                 |  | |  |  |     ||                    |   ||:|:
Db         67 GGGGSGNRNGRGGGGGSGGGGGG----------------RDDCWSEEATKVLI------ 103

Qy        139 EEWGQQQRQLGKKKHRRRPSKKKRHWKPYYKLTWEE------------------KKKFD 179
              | || :  : ||       : |::|||   ::  :                   |||:
Db        104 EAWGDRFSEPGK------GTLKQQHWKEVAEIVNKSRQCKYPKTDIQCKNRIDTVKKKYK 157

Qy        180 EKQSLRAS----------------------------------------RIRAEMF---AK 196
              ::::  ||                                        |: ||    |
Db        158 QEKAKIASGDGPSKWVFFKKLESLIGGTTTFIASSKASEKAPMGGALGNSRSSMFKRQTK 217

Qy        197 GQPVAPYNTTQFLMDDHDQEEPDLKTGLYSKRAAAKSDDTSDDD---FMEEGGE------ 247
              |        | :   ::   |     : ||:|::::  || :     ||  |
Db        218 G--------NQIVQQQQEKRGSDSMRWHFRKRSASETESESDPEPEASPEESAESLPPLQ 269

Qy        248 ----------------EDGGSDGMGGDGSEFLQRDFSETYERYHTESLQNMSKQELIKE 290
                           ||  |:|      |   |:| ||:  |  |: |:  || ||
Db        270 PIQPLSFHMPKRLKVDKSGGGGSGVGDVARAIL--GFTEAYEKAETAKLKLMA--ELEKE 325

Qy        291 YLEL--EKCLSRM--------------EDENNRLRLESKRLGGDDAR 321
               ::   |  | ||              |:| :| | | : :  || |
Db        326 RMKFAKEMELQRMQFLKTQLEITQNNQEEEERSRQRGERRIVDDDDDR 373
```

Search completed: February 5, 2004, 13:41:15
Job time : 93 secs